SPSS TUTORIAL HANDBOOK

ဒေါက်တာနေဝင်းအောင် (ဖွံ့ဖြိုးမှုပညာ)

SPSS တွင် အသုံးများ Statistical Test များ ပြုလုပ်နည်း

၂၀၂၆

မာတိကာ

Chapter ၁ Introduction.....၁

Chapter ၂ TRANSFORM / SELECT DATA (ဒေတာများကို ပြောင်းလဲခြင်းနှင့် ရွေးချယ်ခြင်း) ၅

Chapter ၃ GRAPHS (ပုံကားချပ်များ) ၉

Chapter ၄ CENTRAL TENDENCIES, MEASURES OF DISTRIBUTION, MEASURES OF ASYMMETRY.....၁၈

Chapter ၅. ESTIMATION AND HYPOTHESIS TESTING (ခန့်မှန်းခြင်းနှင့် Hypothesis စမ်းသပ်ခြင်း).....၂၂

Chapter ၆– STATISTICAL DEPENDENCE (ကိန်းဂဏာန်းဆိုင်ရာ ဆက်နွယ်မှုများ).....၂၈

Chapter ၇. CORRELATION AND LINEAR REGRESSION (ဆက်စပ်မှုနှင့် မျဉ်းဖြောင့်ပုံစံ ခန့်မှန်းခြင်း) ၃၄

Chapter ၈. MULTIPLE CORRELATION AND LINEAR REGRESSION (ဆက်စပ်မှုအမြောက်အများနှင့် မျဉ်းဖြောင့်ပုံစံ ခန့်မှန်းခြင်း) ၄၀

Chapter–၉. CURVILINEAR REGRESSION (ကွေးညွှတ်သော မျဉ်းပုံစံဖြင့် ခန့်မှန်းတွက်ချက်ခြင်း) ၄၄

Chapter ၁၀: TIME SERIES ANALYZES (အချိန်နှင့်အမျှ ပြောင်းလဲမှုများကို ဆန်းစစ်ခြင်း) ၄၈

Chapter 1 Introduction

Introduction to SPSS (SPSS အကြောင်း မိတ်ဆက်)

- ပထမဆုံးအနေနဲ့ Csernely_data.sav ဆိုတဲ့ file ကို အရင်ဖွင့်လိုက်ပါ။

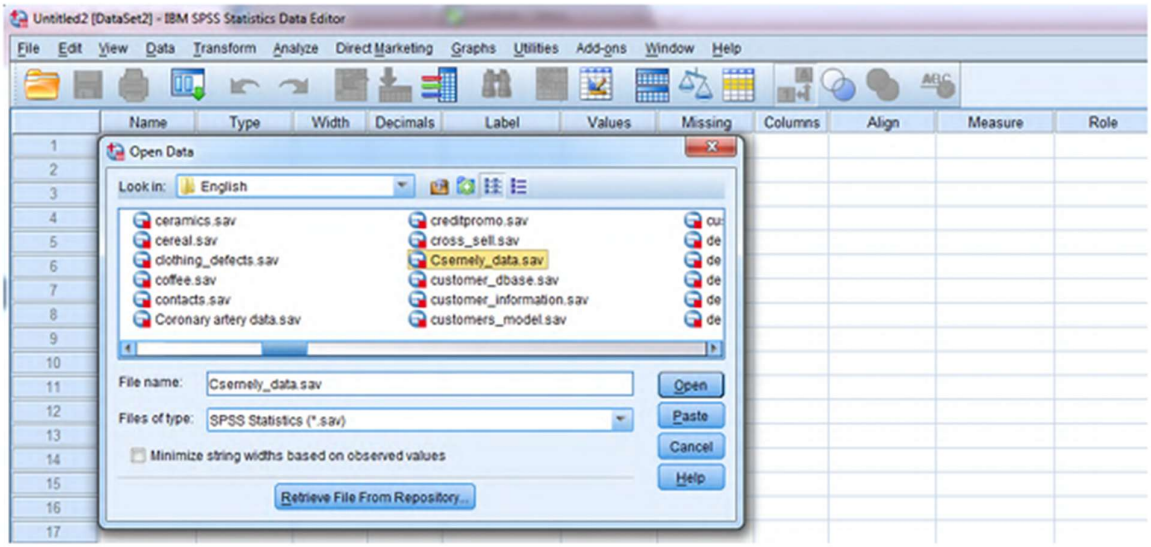


Figure 1 – Opening an SPSS Data File

- Data Editor မှာ data တွေကို ကြည့်လို့ရတဲ့ view ၂ မျိုးရှိပါတယ်၊ အဲဒါတွေကတော့ Data View နဲ့ Variable View တို့ပဲ ဖြစ်ပါတယ်။
- ဒီ view နှစ်ခုစလုံးမှာ အချက်အလက်တွေကို ပြင်တာ (edit) ဒါမှမဟုတ် ဖျက်တာ (delete) တွေ လုပ်လို့ရပါတယ်။
- Data View ကတော့ လက်တွေ့ data တန်ဖိုးတွေ ဒါမှမဟုတ် value labels တွေကို ပြပေးတာပါ။

Variable View (ပြောင်းလဲကိန်းများအား ကြည့်ရှုစစ်ဆေးခြင်း)

Variable View မှာတော့ variable တစ်ခုချင်းစီရဲ့ သတ်မှတ်ချက်တွေကို အခုလို တွေ့ရပါမယ် -

- **Name:** ဒါက variable တစ်ခုချင်းစီရဲ့ သီးသန့်အမည်ဖြစ်ပြီး တစ်ခုနဲ့တစ်ခု မတူရပါဘူး။ အမည်ပေးတဲ့အခါ space ခြားတာမျိုး ဒါမှမဟုတ် တခြား character တွေ သုံးလို့မရသလို၊ ပထမဆုံးစာလုံးကတော့ letter ဖြစ်ရပါမယ်။
- **Type:** ဒါက variable တစ်ခုချင်းစီမှာရှိတဲ့ data အမျိုးအစားကို သတ်မှတ်ပေးတာပါ။ မူလ setting ကတော့ အသုံးအများဆုံးဖြစ်တဲ့ numeric type ဖြစ်ပြီး ကိန်းဂဏန်း variable

တွေကို ဆိုလိုတာပါ။ ဒါပေမဲ့ လိုအပ်ရင် Comma, Dot, Scientific Notation, Date, Dollar, Custom Currency ဒါမှမဟုတ် String အဖြစ် ပြောင်းလို့ရပါတယ်။

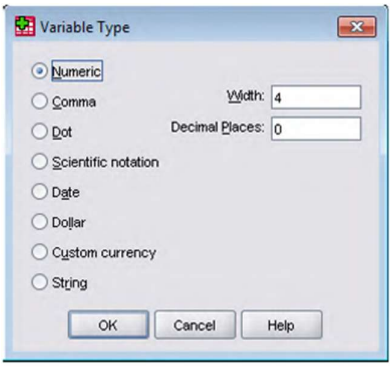


Figure 2 – Variable Type Dialog Box

- **Width:** ဒါက field ရဲ့ အကျယ်ကို ပြောတာပါ။
- **Decimals:** ဒါက numeric type အတွက် ဒသမ ဘယ်နှနေရာ ထားမလဲဆိုတာ သတ်မှတ်တာပါ။
- **Label:** ဒါက variable ရဲ့ အမည်ကို ပိုပြီး အသေးစိတ် ဖော်ပြတာပါ (စာလုံးရေ ၂၅၆ လုံးအထိ သုံးလို့ရပါတယ်)။ သူ့ကိုတော့ Name မှာ သုံးလို့မရတဲ့ space တွေ ဒါမှမဟုတ် တခြား character တွေနဲ့ ရေးလို့ရပါတယ်။
- **Values:** ဒီနေရာမှာ variable တစ်ခုချင်းစီရဲ့ တန်ဖိုးတွေအတွက် အသေးစိတ် value labels တွေ ပေးလို့ရပါတယ်၊ အဲဒီအခါ numeric codes တွေက non-numeric categories တွေကို ကိုယ်စားပြုသွားတာပေါ့။

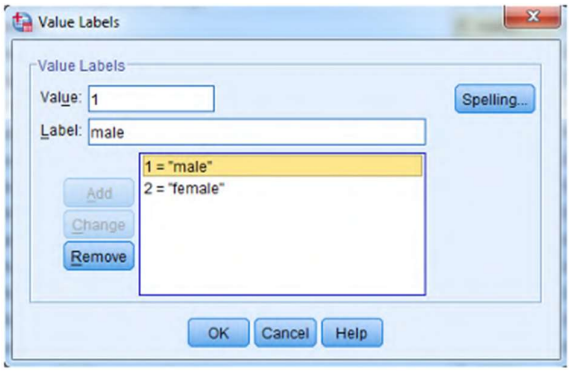


Figure 3 – Value Labels Dialog Box

- **Missing:** ဒါက အချက်အလက် မရှိတဲ့အခါ (ဥပမာ- ဖြေဆိုသူက အဖြေပေးဖို့ ငြင်းဆန်တာမျိုး) မှာ သုံးတာပါ။ user-missing values တွေကို သီးသန့်သတ်မှတ်ထားနိုင်ပြီး တွက်ချက်မှု အများစုမှာ ထည့်မတွက်ဘဲ ချန်လှပ်ထားမှာဖြစ်ပါတယ်။

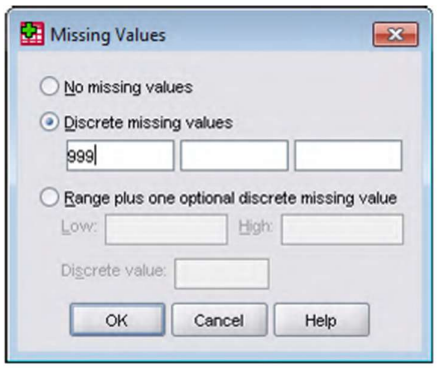


Figure 4 – Missing Values Dialog Box

- **Column:** ဒါက column အကျယ်အတွက် စာလုံးအရေအတွက်ကို သတ်မှတ်တာပါ။
- **Align:** ဒါက data တွေကို display ပြတဲ့နေရာမှာ right, left ဒါမှမဟုတ် center ဆိုပြီး နေရာညှိတာပါ။
- **Measures:** ဒါက scales of measurement တွေကို ပြောတာဖြစ်ပြီး nominal, ordinal, interval ဒါမှမဟုတ် ratio scale တွေ ဖြစ်နိုင်ပါတယ်။ SPSS ထဲမှာတော့ nominal, ordinal နဲ့ ratio measures ဆိုပြီး တွေ့ရပါလိမ့်မယ်။

Scales of Measurement (တိုင်းတာမှု အဆင့်အတန်းများ)

- **Nominal scale:** ကိန်းဂဏန်းတွေကို labels တွေ၊ အုပ်စုတွေ ဒါမှမဟုတ် classes တွေအနေနဲ့ပဲ သုံးတာပါ။ အရာဝတ္ထုတွေကို labels အဖြစ် သတ်မှတ်ပေးတဲ့ ရိုးရှင်းတဲ့ codes တွေပေါ့။ သူ့ကို qualitative data တွေ (ဥပမာ- အလုပ်အကိုင်ခွဲခြားတာ၊ ပထဝီဝင်အရ ခွဲခြားတာ) အတွက် သုံးပါတယ်။ ဥပမာ- blonde ဆိုရင် ၁၊ brown ဆိုရင် ၂၊ red ဆိုရင် ၃၊ black ဆိုရင် ၄ လို့ သတ်မှတ်တာမျိုးပါ။ ဆံပင်နီ (၃) တဲ့သူက blonde (၁) တဲ့သူထက် ပိုပြီး 'ဆံပင်များတယ်' လို့ ပြောလို့မရပါဘူး။
- **Ordinal scale:** data တွေကို သူတို့ရဲ့ အရွယ်အစား ဒါမှမဟုတ် အရည်အသွေးအလိုက် အစဉ်လိုက် စီလို့ရပါတယ်၊ ပေးထားတဲ့ ကိန်းဂဏန်းတွေက rank order (1st, 2nd, 3rd စသဖြင့်) ကို ကိုယ်စားပြုပါတယ်။ ဥပမာ- ကုမ္ပဏီတွေရဲ့ top lists တွေလိုမျိုးပေါ့။
- **Interval scale:** တိုင်းတာချက် နှစ်ခုကြားက distances က အဓိပ္ပာယ်ရှိပါတယ်။ ဒါပေမဲ့ zero point ကတော့ စိတ်ကြိုက်သတ်မှတ်တာဖြစ်ပြီး negative values တွေလည်း သုံးလို့ရပါတယ်။ ဒီ scale ပေါ်က ကိန်းဂဏန်းတွေကြားက ratios တွေက အဓိပ္ပာယ်မရှိတဲ့အတွက် multiplication နဲ့ division တွေကို တိုက်ရိုက်လုပ်လို့မရပါဘူး။ ဥပမာ- Celsius scale နဲ့ တိုင်းတဲ့ temperature လိုမျိုးပေါ့။

- **Ratio scale (Scale):** ဒါက အားအကောင်းဆုံး scale of measurement ပါ။ တိုင်းတာချက်တွေကြားက distances ရော ratios ရောက အဓိပ္ပာယ်ရှိသလို၊ အဓိပ္ပာယ်ရှိတဲ့ zero point လည်း ပါဝင်ပါတယ်။ ဥပမာ- mass, length တို့လိုမျိုး ဖြစ်ပါတယ်။

Chapter 2 TRANSFORM / SELECT DATA (ဒေတာများကို ပြောင်းလဲခြင်းနှင့် ရွေးချယ်ခြင်း)

Example 1: ဖြေဆိုသူတွေရဲ့ အသက်က ဘယ်လောက်လဲ။ 'age' ဆိုတဲ့ variable အသစ်တစ်ခု တည်ဆောက်ကြည့်ရအောင်။

- ဒေတာတစ်ခုကို ပြောင်းလဲပြီး (transform လုပ်ပြီး) variable အသစ်တွေ တည်ဆောက်လို့ ရပါတယ်။
- ကျွန်တော်တို့ဆီမှာ ဝန်ထမ်းတွေရဲ့ မွေးသက္ကရာဇ် (date of birth) ဒေတာရှိတယ်ဆိုပါစို့။ လက်ရှိနှစ်ထဲကနေ မွေးသက္ကရာဇ်ကို နှုတ်လိုက်ရင် သူတို့ရဲ့ အသက်ကို ရပါလိမ့်မယ်။
- အသက်ကို တွက်ချက်ဖို့အတွက် menu ထဲက **Transform / Compute Variable...** ကို ရွေးချယ်ပါ။

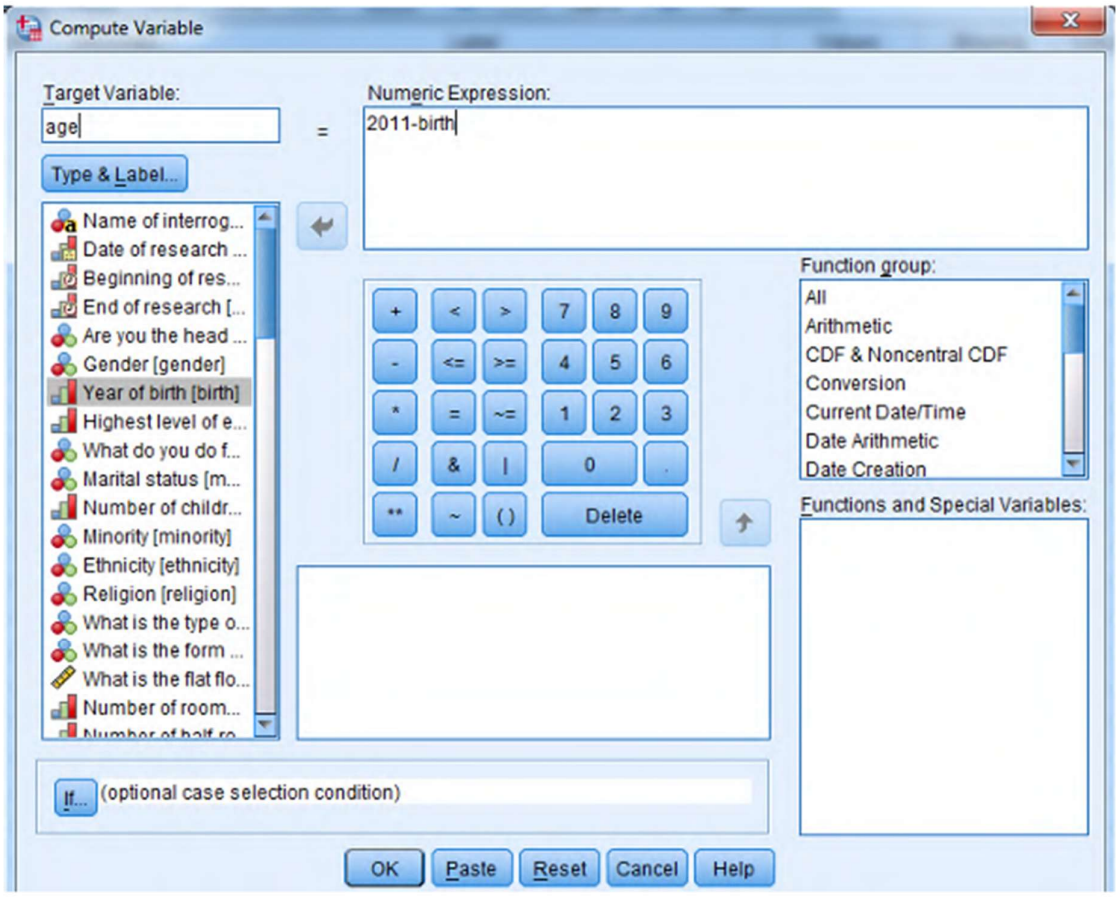


Figure 5 – Compute Variable Dialog Box

- Target variable နေရာမှာ မိမိပေးချင်တဲ့ အမည် (ဥပမာ - age) ကို ရိုက်ထည့်ပါ။

- Expression field ထဲမှာတော့ တွက်ချက်မယ့် ပုံသေနည်း အစိတ်အပိုင်းတွေကို ရိုက်ထည့်ရမှာပါ။ အကယ်၍ မွေးသက္ကရာဇ်က Date (mm/dd/yyyy) ပုံစံနဲ့ ရှိနေရင်တော့ အဲဒီထဲကမှ year အပိုင်းကိုပဲ ထုတ်ယူဖို့ လိုပါလိမ့်မယ်။
- ပုံသေနည်း ရေးပြီးသွားရင် OK ကို နှိပ်လိုက်တာနဲ့ variable အသစ်တစ်ခု အသင့်ဖြစ်သွားပါပြီ။

Example 2: လူပျို/အပျို (single) ဖြစ်သူတွေရဲ့ အချိုးအစား (proportion) က ဘယ်လောက်လဲ။

- menu ထဲက **Analyze / Descriptive Statistics / Frequencies…** ကို ရွေးပါ။
- Relative frequency တွက်ချက်ချင်တဲ့ variable (Marital status) ကို ရွေးပြီး OK နှိပ်ပါ။

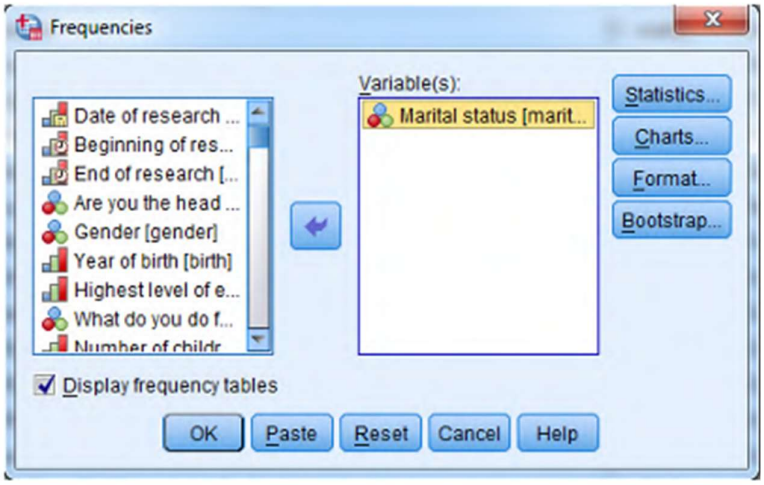


Figure 6 – Frequencies Dialog Box

- ရလာတဲ့ ရလဒ်ကို Output View မှာ ကြည့်နိုင်ပါတယ်။

Marital status

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	single	12	**5.4**	**5.5**	5.5
	married	102	45.9	46.6	52.1
	divorced	27	12.2	12.3	64.4
	partner (but not married)	21	9.5	9.6	74.0
	widow	57	25.7	26.0	100.0
	Total	219	98.6	100.0	
Missing	System	3	1.4		
Total		222	100.0		

- Frequency table အရ Csernely မှာရှိတဲ့ လူတွေရဲ့ ၅.၄ ရာခိုင်နှုန်းက လူပျို/အပျိုတွေ ဖြစ်ပါတယ် (ဖြေဆိုသူတွေထဲကဆိုရင်တော့ ၅.၅ ရာခိုင်နှုန်းပေါ့)။

Example 3: ပင်စင်စား (pensioners) တွေထဲမှာ အမျိုးသား ရာခိုင်နှုန်း ဘယ်လောက်ရှိသလဲ။

- ဒီတစ်ခါမှာတော့ statistical population က ဖြေဆိုသူ အားလုံးမဟုတ်တော့ဘဲ ပင်စင်စားတွေပဲ ဖြစ်ပါတယ်။
- ဒါကြောင့် အချက်အလက်တွေကို အရင်ဆုံး subset အနေနဲ့ ရွေးချယ်ဖို့ **Data / Select cases...** ကို သုံးရပါမယ်။

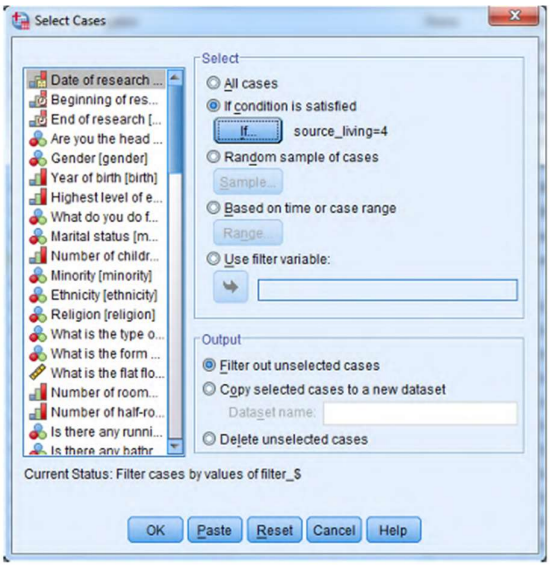


Figure 7 – Select Cases Dialog Box

- အမျိုးသားတွေကိုပဲ သီးသန့်ရွေးချယ်ဖို့အတွက် conditional expression ဖြစ်တဲ့ **Gender = 1** (အမျိုးသားအတွက် code ကို ၁ လို့ သတ်မှတ်ထားလို့ပါ) ကို သုံးရပါမယ်။

	jobcat
1	unemployment benefit
2	.
3	pensioner
4	pensioner

Figure 8 – Selected Subset of Cases

- ပြီးရင် **Analyze / Descriptive Statistics / Frequencies...** ကို ထပ်သွားပါ။ ကျား/မ ရာခိုင်နှုန်းကို သိချင်တာဖြစ်တဲ့အတွက် Gender ကို Variables ထဲ ထည့်လိုက်ပါ။

Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	62	**50.4**	50.4	50.4
	Female	61	49.6	49.6	100
	Total	123	100.0	100.0	

- Output view မှာ ထွက်လာတဲ့ ရလဒ်အရ ပင်စင်စားတွေရဲ့ ၅၀.၄ ရာခိုင်နှုန်းက အမျိုးသားတွေ ဖြစ်ပါတယ်။

Chapter 3 GRAPHS (ပုံကားချပ်များ)

Example 1: ဖြေဆိုသူတွေရဲ့ ကျား/မ အချိုးအစားကို bar chart ဆွဲကြည့်ကြရအောင်။ ပုံကို ပိုလှအောင်ပြင်ပြီး အချိုးအစား တန်ဖိုးတွေကိုပါ ပုံမှာ ထည့်သွင်းပြသရပါမယ်။

- bar chart ဆွဲဖို့အတွက် menu ထဲက **Graphs / Legacy Dialogs / Bar Charts / Simple** ကို ရွေးပေးပါ။

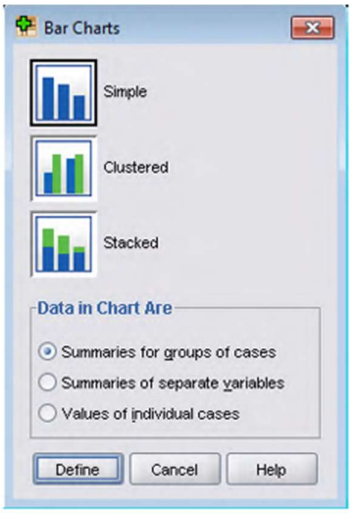


Figure 9 – Bar Charts Dialog Box

- ပြီးရင် Category axis နေရာမှာ variable (gender) ကို ထည့်ပါ။ ကျွန်တော်တို့က ဖြေဆိုသူတွေရဲ့ အချိုးအစားကို သိချင်တာဖြစ်တဲ့အတွက် Bars represent နေရာမှာ **% of cases** ကို ရွေးပေးရပါမယ်။

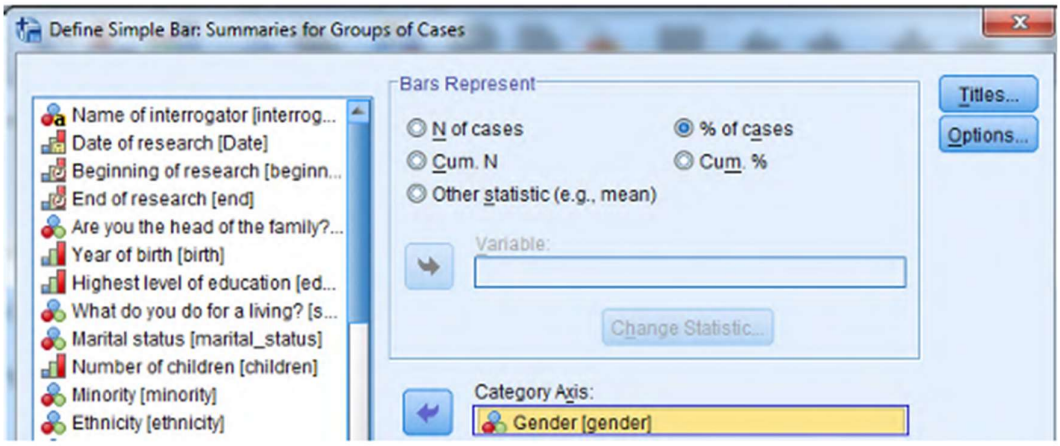


Figure 10 – Bar Charts Dialog Box

- ရလာတဲ့ ပုံကို Output window မှာ တွေ့ရပါလိမ့်မယ်။ ပုံကို ပြင်ချင်ရင်တော့ အဲဒီပုံပေါ်မှာ double click နှိပ်လိုက်ရင် Chart Editor window ပွင့်လာပါမယ်။ (ဒါမှမဟုတ် ပုံပေါ်မှာ right-click နှိပ်ပြီး Edit Content ထဲက 'In separate window' ကို ရွေးလို့လည်း ရပါတယ်)။
- အဲဒီမှာ ပုံရဲ့ အရောင်၊ depth နဲ့ angle (3D)၊ bars တွေရဲ့ width စတာတွေကို စိတ်ကြိုက် ပြင်လို့ရပါတယ်။ အချိုးအစား တန်ဖိုးတွေကို ပုံမှာ ပေါ်စေချင်ရင်တော့ **Data Label Mode** ကို ရွေးပေးပါ။

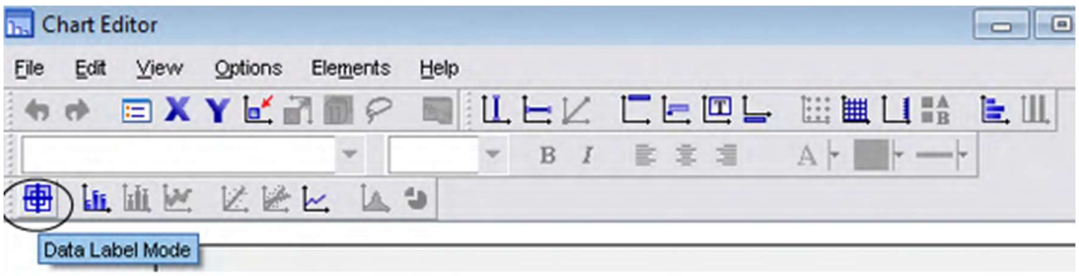


Figure 11 – Chart Editor

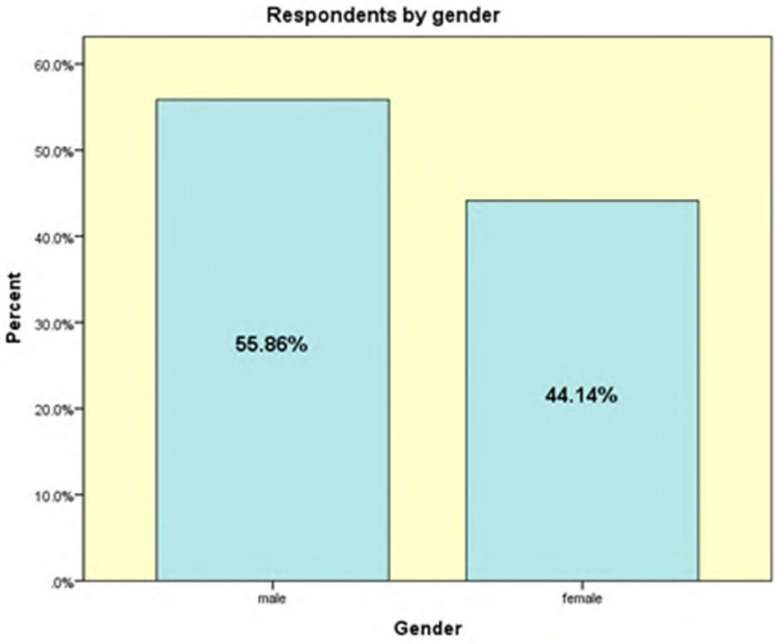


Figure 12 – Bar Chart

Example 2: ခုနက ဆွဲထားတဲ့ bar chart ကို pie chart အဖြစ် ပြောင်းကြည့်ရအောင်။

- ပုံစံပြောင်းဖို့အတွက် Chart Editor ထဲက bar chart ကို နှိပ်ပြီး menu ကနေ **Edit / Properties / Variables / Element Type...** ထဲမှာ Pie ကို ရွေးပေးလိုက်ရုံပါပဲ။

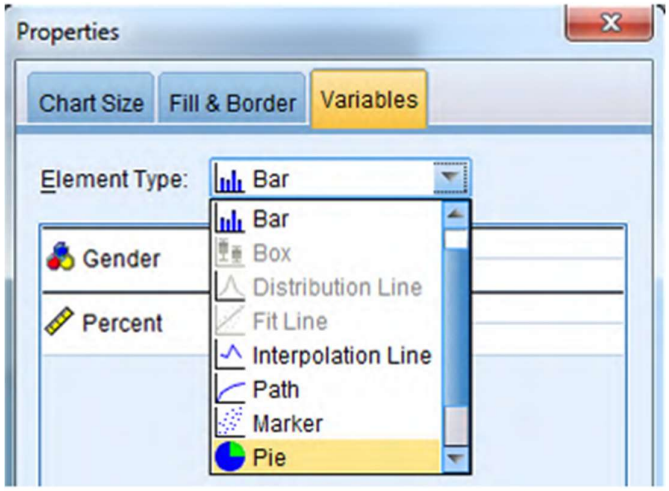


Figure 13 – Chart Editor Properties Dialog Box

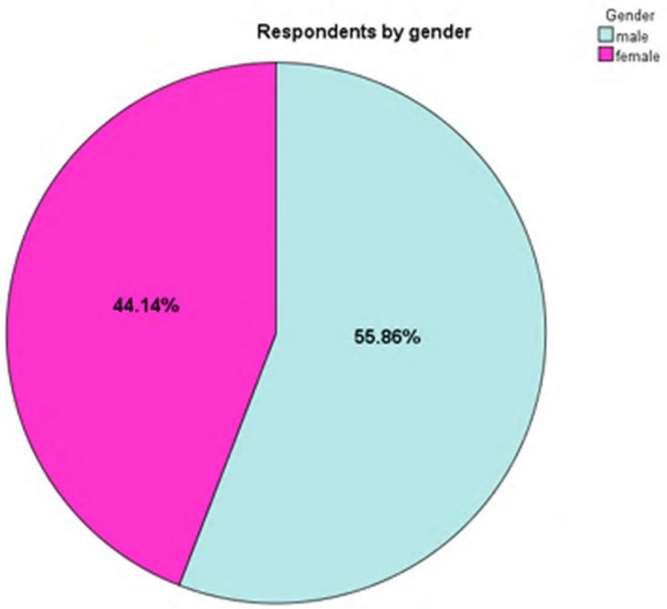


Figure 14 – Pie Chart

Example 3: ဖြေဆိုသူတွေရဲ့ ပညာအရည်အချင်းအလိုက် အချိုးအစားကို ကျား/မ ခွဲပြီး stacked bar chart ဆွဲကြည့်ရအောင်။

- ဒီနေရာမှာ Example 1 နဲ့ ကွာတာက **stacked bar chart** ကို သုံးရမှာပါ။

- Graphs menu ထဲက Legacy Dialogs / Bar Charts ကိုသွားပြီး stacked အမျိုးအစားကို ရွေးပါ။ ပြီးရင် Category axis မှာ ပညာအရည်အချင်း (Education level) ကို ထည့်ပြီး Stacks နေရာမှာ ကျား/မ (Gender) ကို ထည့်ပေးပါ။

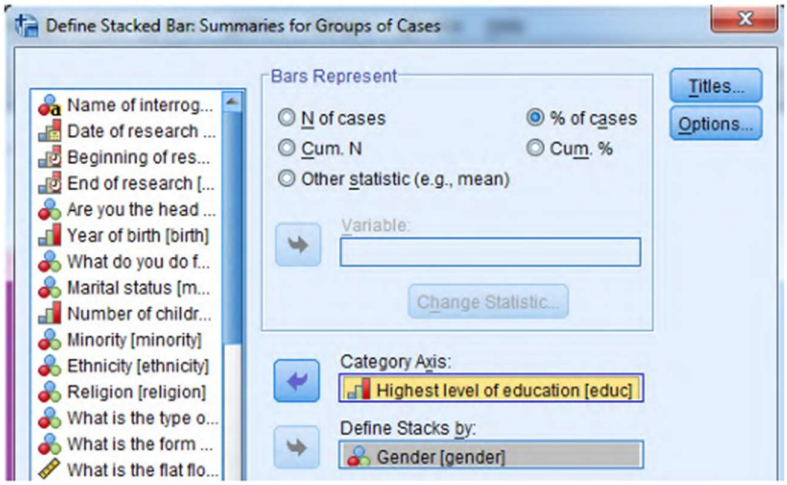


Figure 15 – Stacked Bar Chart Dialog Box

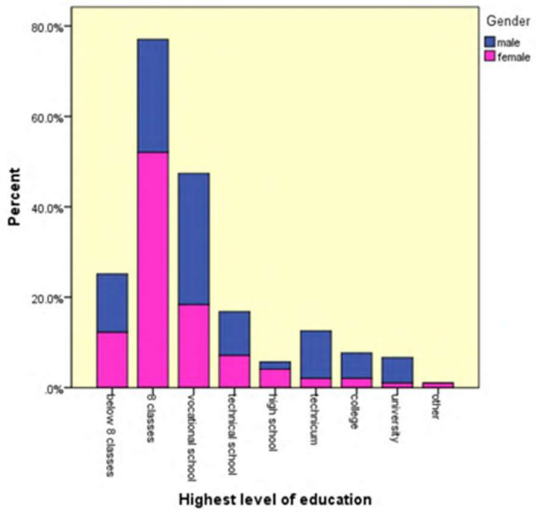


Figure 16 – Stacked Bar Chart

Example 4: အိမ်ထောင်စုတွေရဲ့ ပျမ်းမျှဝင်ငွေနဲ့ စုစုပေါင်းအသုံးစရိတ်ကြား ဆက်စပ်မှုကို အပူပေးစနစ် (heating_bin) အလိုက် ခွဲပြီး scatter plot ဆွဲကြည့်ရအောင်။

- scatter plot ဆွဲဖို့ **Graphs / Legacy Dialogs / Scatter/Dot...** ကို နှိပ်ပြီး Simple scatter ကို ရွေးပါ။

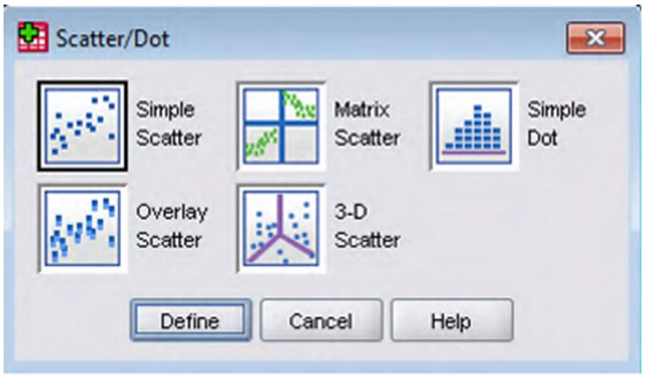


Figure 17 – Scatter / Dot Dialog Box

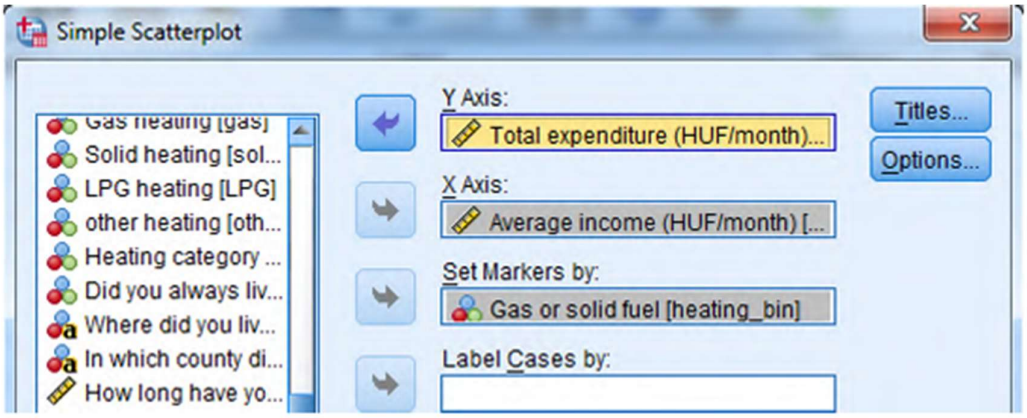


Figure 18 – Simple Scatter Plot Dialog Box

- အရင်ဆုံး ဝင်ရိုးတွေကို သတ်မှတ်ပေးရပါမယ် (x: average income; y: total expenditure)။ ပြီးရင် Set markers by နေရာမှာ အပူပေးစနစ် (heating_bin) ကို ထည့်ပါ။
- လိုအပ်ရင် Label Cases by မှာ variable တစ်ခုခုနဲ့ အမည်တပ်လို့ ရပါတယ်။ Chart Editor ကို သုံးပြီး အစက်အပြောက်တွေရဲ့ အရောင်နဲ့ ပုံစံတွေကိုလည်း ပြင်နိုင်ပါတယ်။

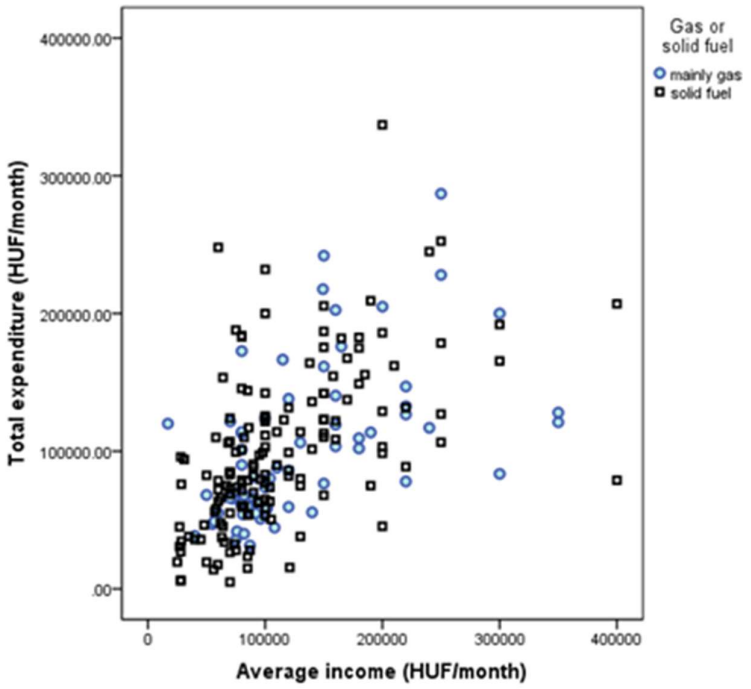


Figure 19 – Simple Scatter Plot

Example 5: စုစုပေါင်းအသုံးစရိတ်ကို horizontal box plot ဆွဲကြည့်ရအောင်။

- အရင်ဆုံး box plot ဆိုတာ ဘာလဲဆိုတာ သိဖို့လိုပါတယ်။ box plot ဆိုတာ median, lower quartile, upper quartile နဲ့ အနည်းဆုံး၊ အများဆုံး တန်ဖိုးတွေလိုမျိုး distribution ရဲ့ အနှစ်ချုပ် တိုင်းတာချက်တွေကို စုစည်းပြသပေးတာပါ။ ဒါ့အပြင် asymmetry ကိုလည်း မြင်နိုင်ပါတယ်။

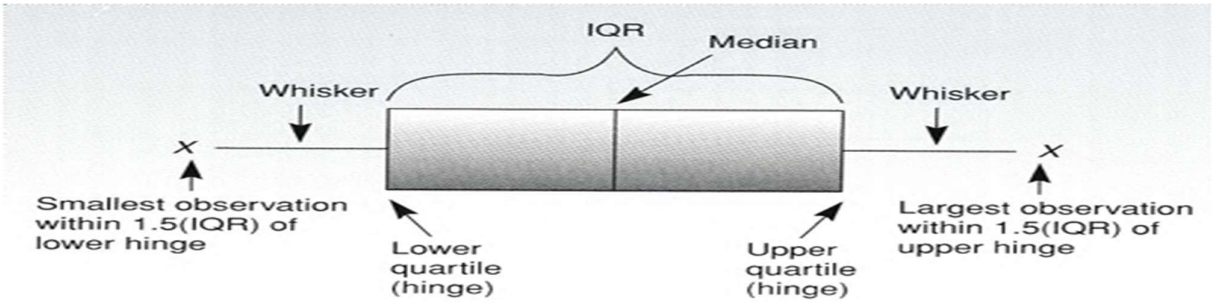


Figure 20 – Box Plot
Source: Aczel, 1996

- box plot ဆွဲဖို့ **Graphs / Legacy Dialogs / Boxplot** ကို သွားပါ။ Variable တစ်ခုတည်းအတွက်ဆိုရင် Simple ကို ရွေးပါ။ ဒေတာတွေက Summaries of separate variables ဖြစ်ရပါမယ်။

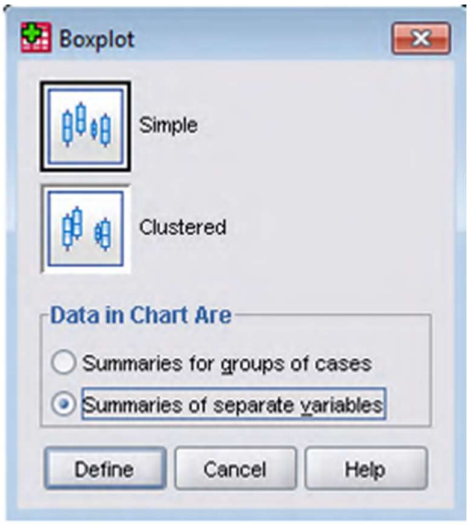


Figure 21 – Box Plot Dialog Box

- ပုံကို အလဲ့ (horizontal) ဖြစ်စေချင်ရင်တော့ Chart Editor ပွင့်လာတဲ့အခါ **Options / Transpose Chart** ကို ရွေးပေးရပါမယ်။

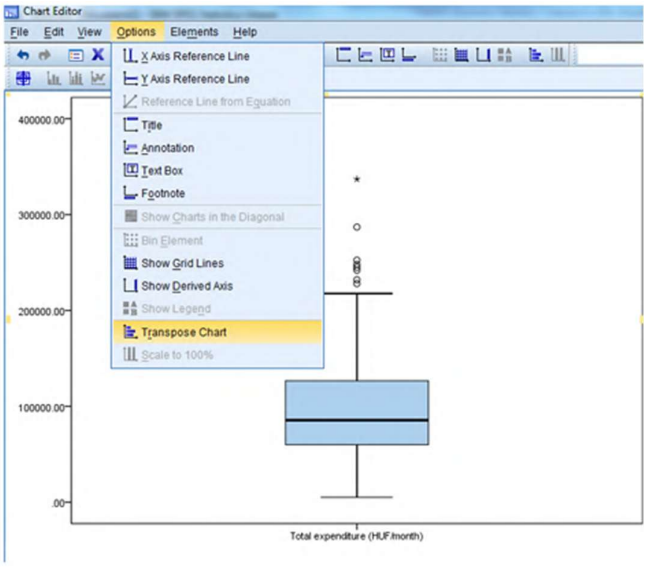


Figure 22 – Box Plot Chart Editor

Example 6: အိမ်ထောင်စုတွေရဲ့ စုစုပေါင်းအသုံးစရိတ်ကို အပူပေးစနစ် (heating_bin) အလိုက် ခွဲခြားပြီး clustered box plot ဆွဲကြည့်ရအောင်။

- ဒီနေရာမှာတော့ clustered box plot ကို ရွေးရပါမယ်။ ဒေတာတွေက Summaries for groups of cases ပုံစံ ဖြစ်ရပါမယ်။
- Variable နေရာမှာ total expenditure ကို ထည့်ပြီး Category axis မှာ heating_bin ကို ထည့်ပါ။ Clusters တွေကိုတော့ household clusters နဲ့ သတ်မှတ်ပေးရပါမယ်။

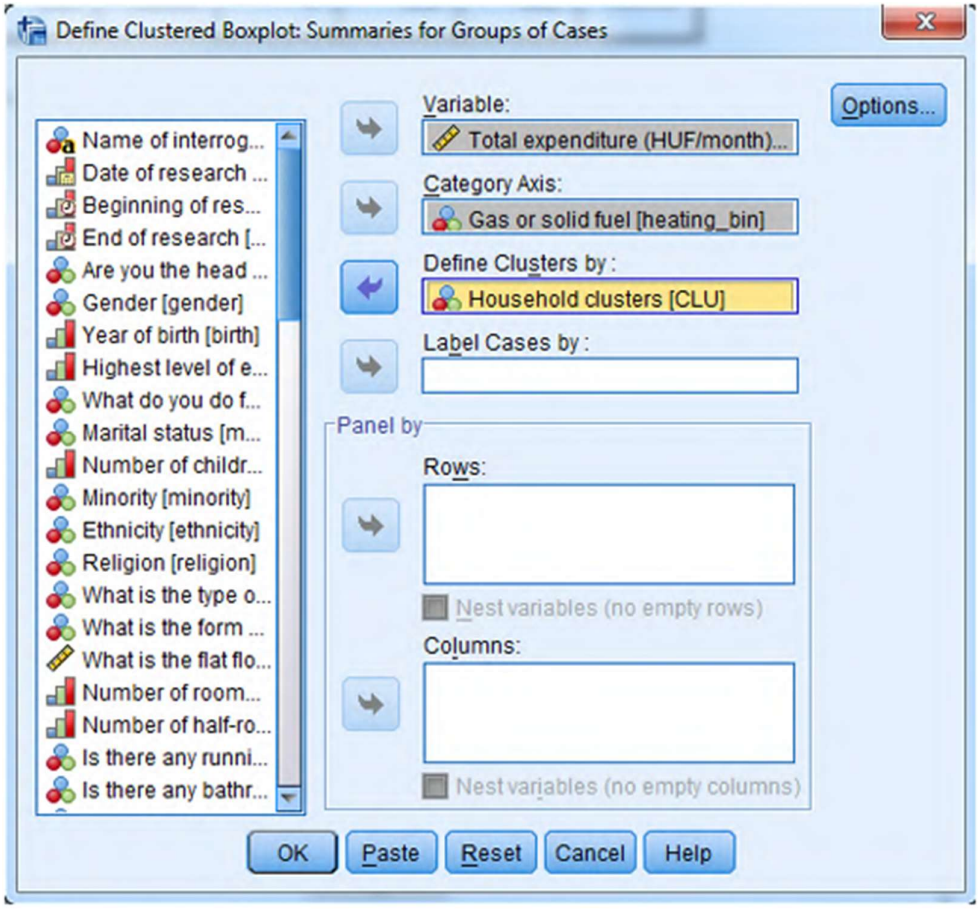


Figure 23 – Clustered Box Plot Chart Editor

- ပုံကို ပြင်ချင်ရင် ပုံပေါ်မှာ double click နှိပ်၊ ပြီးရင် မိမိပြင်ချင်တဲ့ အစိတ်အပိုင်းပေါ်မှာ double click ထပ်နှိပ်ရင် Properties dialog box ကျလာပါလိမ့်မယ်။ ပြင်ပြီးရင် Apply ကိုနှိပ်ပြီး dialog box ကို ပိတ်လိုက်ပါ။

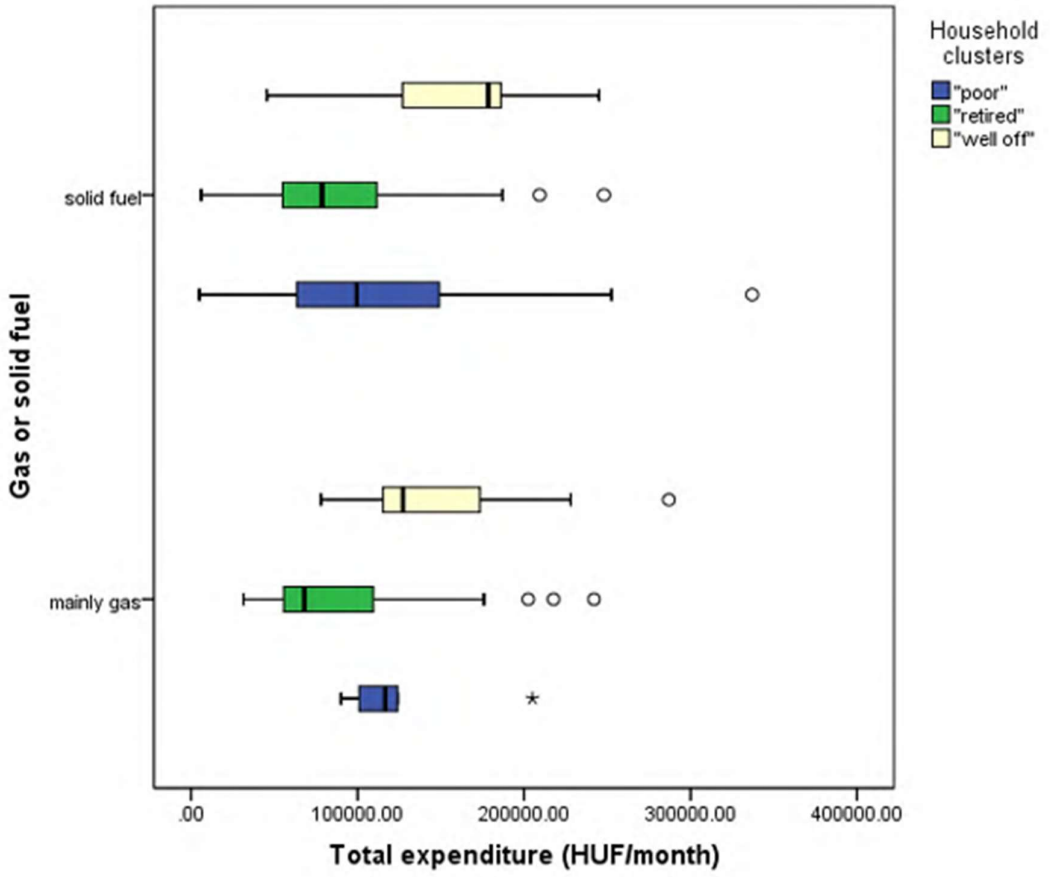


Figure 24 – Clustered Box Plot

Chapter 4 CENTRAL TENDENCIES, MEASURES OF DISTRIBUTION, MEASURES OF ASYMMETRY

ဥပမာ - အိမ်ထောင်စုတွေရဲ့ တစ်လတာ စုစုပေါင်း အသုံးစရိတ် (HUF/month) အတွက် central tendencies, measures of distribution, measures of asymmetry နဲ့ quartiles တွေကို ရှာကြည့်ရအောင်။

- ဒီအချက်အလက်တွေက descriptive statistics တိုင်းတာချက်တွေ ဖြစ်ပြီး **Analyze menu** ထဲက **Descriptive Statistics / Descriptive…** ကို နှိပ်ပြီး ရယူနိုင်ပါတယ်။
- ပြီးရင် ကိုယ်စစ်ဆေးချင်တဲ့ variable ကို ရွေးပေးရပါမယ်။
- အကယ်၍ လိုချင်တယ်ဆိုရင် variable တွေကို standardize လုပ်လို့လည်း ရပါတယ်။

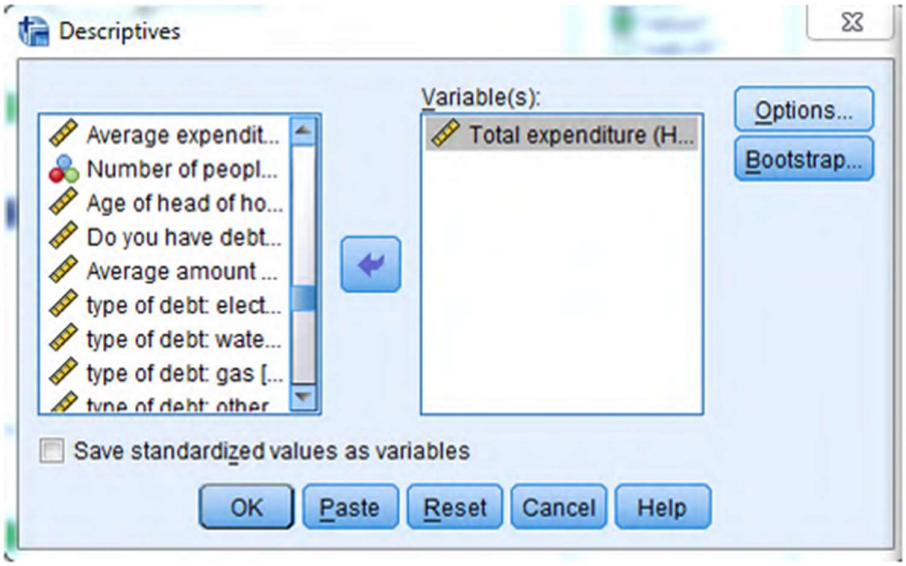


Figure 25 – Descriptives Dialog Box

- Statistics တွေကို ပိုပြီး စုံစုံလင်လင် သိချင်ရင်တော့ **Options** ကို နှိပ်ပြီး ရွေးချယ်နိုင်ပါတယ်။

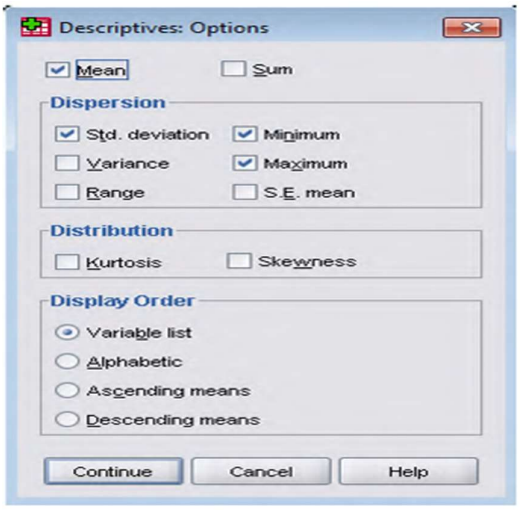


Figure 26 – Clustered Box Plot

- ဒီ option ကို သုံးပြီး mean, standard deviation နဲ့ asymmetry တိုင်းတာချက်တွေကို သတ်မှတ်လို့ ရပါတယ်။
- နောက်တစ်နည်းကတော့ **Analyze / Descriptive Statistics / Frequencies...** menu ကနေတစ်ဆင့် သွားတာပါ။

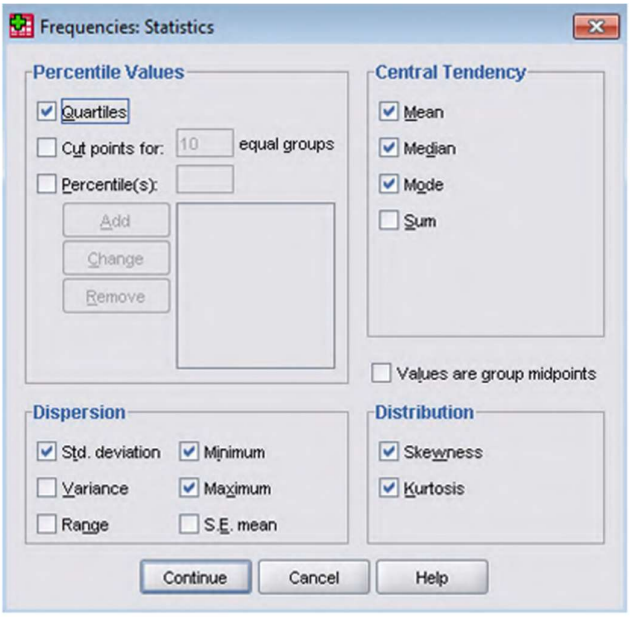


Figure 27 – Frequencies Statistics Dialog Box

- ပြီးရင် **Statistics** ကို နှိပ်လိုက်ရင် လိုအပ်တဲ့ box ကျလာပါလိမ့်မယ်။

Statistics

Total expenditure (HUF/month)

N	Valid	222
	Missing	0
Mean		99692.72
Median		85500.00
Mode		58500.00[a]
Std. Deviation		57045.35
Skewness		1.02
Kurtosis		1.25
Minimum		5000.00
Maximum		337000.00
Percentiles	25	59875.00
	50	85500.00
	75	126700.00

a. Multiple modes exist. The smallest value is shown

- ထွက်လာတဲ့ ရလဒ်တွေကို Output window မှာ တွေ့ရမှာ ဖြစ်ပြီး အဲဒီကိန်းဂဏန်းတွေကို အခုလို အဓိပ္ပာယ်ဖွင့်နိုင်ပါတယ် -
    - Csernely မှာရှိတဲ့ အိမ်ထောင်စုပေါင်း ၂၂၂ စုကို လေ့လာခဲ့တာ ဖြစ်ပါတယ် (Number of cases)။
    - တစ်လတာ ပျမ်းမျှ အသုံးစရိတ်ကတော့ ၉၉,၆၉၂.၇၂ HUF ဖြစ်ပါတယ်။
    - အိမ်ထောင်စု တစ်ဝက်က ၈၅,၅၀၀ HUF ထက် ပိုသုံးပြီး ကျန်တဲ့တစ်ဝက်ကတော့ အဲဒီထက် နည်းနည်းပဲ သုံးကြပါတယ် (median တန်ဖိုး ၈၅,၅၀၀ HUF ရဲ့ အထက်နဲ့ အောက်မှာ case အရေအတွက် တစ်ဝက်စီ ရှိနေတာပါ)။
    - အသုံးအများဆုံး လစဉ် အသုံးစရိတ်ကတော့ ၅၈,၅၀၀ HUF ဖြစ်ပါတယ်။ mode တန်ဖိုး အများကြီး ရှိနေတဲ့အထဲက အငယ်ဆုံး တန်ဖိုးကို ပြထားတာပါ။
    - ပျမ်းမျှကိန်း (mean) ရဲ့ ပတ်လည်မှာ ပြန့်နှံ့နေတဲ့ ပျမ်းမျှ အကွာအဝေး (average dispersion) ကတော့ ၅၇,၀၄၅.၃၅ HUF ရှိပါတယ်။
    - ညာဘက် အမြီးရှည်တဲ့ ပုံစံ (Long right tail asymmetry) ရှိနေပါတယ်။ skewness က အပေါင်းကိန်း သိသိသာသာ ဖြစ်နေရင် ညာဘက်ကို ပိုစောင်းနေတာပါ။
    - Kurtosis က အပေါင်းကိန်း ဖြစ်နေတဲ့အတွက် normal distribution နဲ့ နှိုင်းယှဉ်ရင် data တွေက ဗဟိုမှာ ပိုစုနေပြီး အမြီးပိုင်းမှာလည်း data ပိုများနေတာကို ပြပါတယ်။
    - အိမ်ထောင်စုတွေရဲ့ အနည်းဆုံး အသုံးစရိတ်က ၅,၀၀၀ HUF ဖြစ်ပါတယ်။

- အိမ်ထောင်စုတွေရဲ့ အများဆုံး အသုံးစရိတ်ကတော့ ၃၃၇,၀၀၀ HUF ဖြစ်ပါတယ်။
- အိမ်ထောင်စု ၂၅ ရာခိုင်နှုန်းက လစဉ် အသုံးစရိတ် ၅၉,၈၇၅ HUF ထက် နည်းပြီး ကျန်တဲ့ ၇၅ ရာခိုင်နှုန်းကတော့ ပိုများပါတယ်။
- အိမ်ထောင်စု ၇၅ ရာခိုင်နှုန်းက လစဉ် အသုံးစရိတ် ၁၂၆,၇၀၀ HUF ထက် နည်းပြီး ကျန်တဲ့ ၂၅ ရာခိုင်နှုန်းကတော့ ပိုများပါတယ်။

Chapter 5. ESTIMATION AND HYPOTHESIS TESTING (ခန့်မှန်းခြင်းနှင့် Hypothesis စမ်းသပ်ခြင်း)

Example 1: Total expenditure အတွက် 95% confidence interval ကို ရှာကြည့်ရအောင်။

- confidence interval ကို သတ်မှတ်ဖို့အတွက် SPSS မှာ **Analyze / Compare Means / One Sample T Test…** ကို နှိပ်ပြီး t-test ကို သုံးနိုင်ပါတယ်။

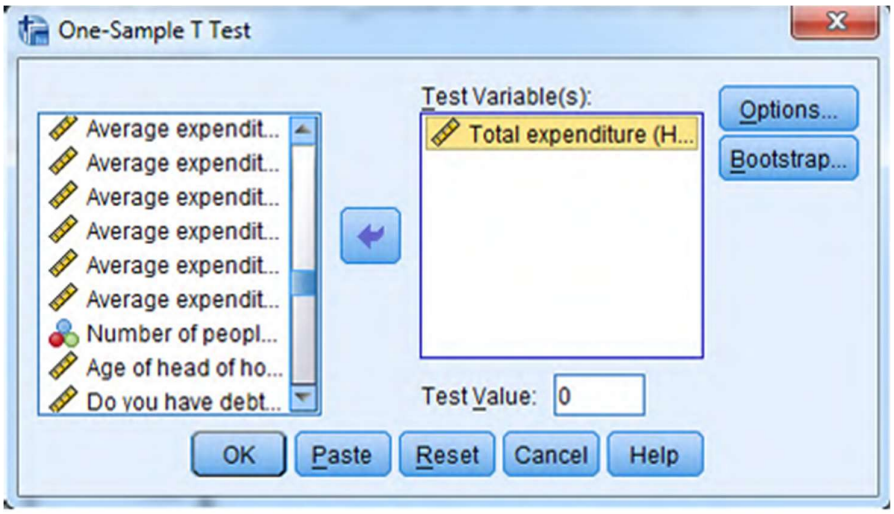


Figure 28 – One-Sample T Test Dialog Box for Estimation

- estimation လုပ်တဲ့အခါမှာတော့ test value ကို သုည (zero) ပဲ ထားရပါမယ်။
- confidence level ကို ပြင်ချင်ရင်တော့ **Options** ကို နှိပ်ပြီး ပြင်နိုင်ပါတယ် (မူလ setting ကတော့ 95% ပါ)။
- OK နှိပ်လိုက်ရင် ရလဒ်တွေက Output window မှာ ပေါ်လာပါလိမ့်မယ်။

Test Value = 0					
t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
				Lower	Upper
26.039	221	0.000	99692.71622	**92147.4136**	**107238.0189**

- ရလာတဲ့ ရလဒ်အရ 95% confidence level မှာ အိမ်ထောင်စုတွေရဲ့ ပျမ်းမျှ total expenditure ဟာ 92 147.4136 နဲ့ 107 238.0189 HUF ကြားမှာ ရှိနေပါတယ်။

Example 2: အိမ်ထောင်စုတွေရဲ့ total expenditure ဟာ $100,000 နဲ့ ညီ၊ မညီ hypothesis ကို စမ်းသပ်ကြည့်ရအောင်။ (alpha = 5%)

- menu ထဲကနေ **Analyze / Compare Means / One Sample T Test…** ကို ရွေးပါ။
- sample mean တစ်ခုချင်းစီကို နှိုင်းယှဉ်ဖို့အတွက် test value နေရာမှာ 100 000 လို့ ရိုက်ထည့်ပါ။

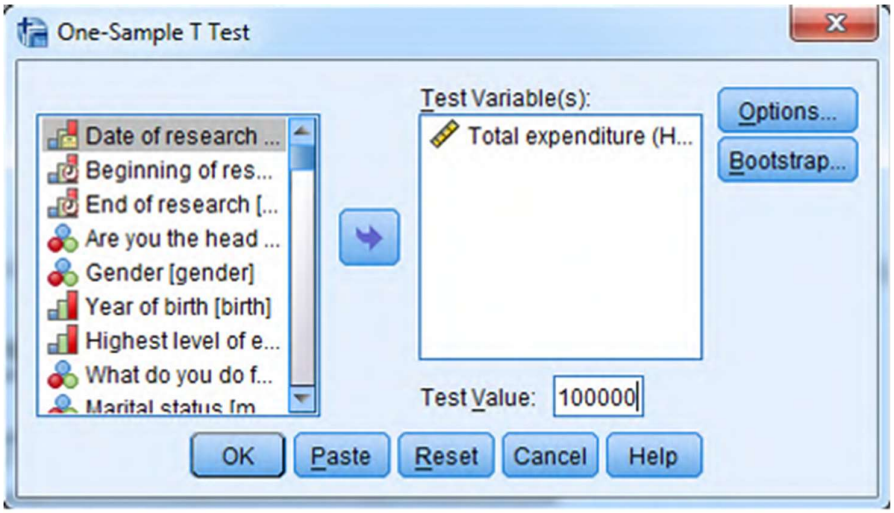


Figure 29 – One-Sample T Test Dialog Box for Hypothesis Testing

- significance level က 5% ဆိုရင် confidence level က 95% ဖြစ်မှာပါ။ ဒါကို Options menu အောက်မှာ ပြင်နိုင်ပါတယ်။

Test Value = 100000					
t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
				Lower	Upper
-.080	221	**.936**	-307.28378	-7852.5864	7238.0189

- p-value (Sig. 2-tailed) က 0.05 ထက် နည်းရင် null hypothesis ကို ပယ်ရမှာဖြစ်ပေမဲ့၊ အခုရလဒ်မှာတော့ p-value က 0.936 ဖြစ်နေတဲ့အတွက် null hypothesis က မှန်ကန်တယ်လို့ ပြောလို့ရပါတယ်။

Example 3: ဓာတ်ငွေ့ (gas) သုံးတဲ့ အိမ်ထောင်စုတွေနဲ့ အခဲလောင်စာ (solid fuels) သုံးတဲ့ အိမ်ထောင်စုတွေရဲ့ ပျမ်းမျှ heating အသုံးစရိတ်ဟာ အတူတူပဲလားဆိုတာ စမ်းသပ်ကြည့်ရအောင်။ (alpha = 5%)

- မသက်ဆိုင်တဲ့ ပျမ်းမျှကိန်း နှစ်ခုကို နှိုင်းယှဉ်ဖို့အတွက် **Analyze / Compare Means / Independent Samples T Test…** ကို သုံးရပါမယ်။

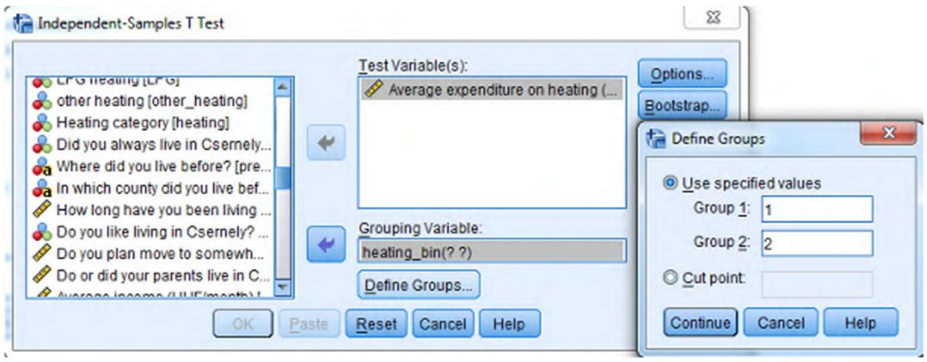


Figure 30 – Independent Samples T Test Dialog Box

- grouping variable နေရာမှာ heating_bin ကို ရွေးပြီး၊ အုပ်စုတွေကိုတော့ gas အတွက် 1 နဲ့ solid fuels အတွက် 2 ဆိုပြီး code တွေ သတ်မှတ်ပေးရပါမယ်။
- variances တွေ ညီ၊ မညီကိုတော့ **Levene's test** နဲ့ စစ်ဆေးပါတယ်။

	Levene's Test for Equality of Variances		t-test for Equality of Means				
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Equal variances assumed	5.012	**0.026**	-1.506	190	**0.134**	-5617.6	3729.59
Equal variances not assumed			-1.716	189.990	0.088	-5617.6	3273.48

- အခု Example မှာ F-value က ကြီးပြီး Levene's test ရဲ့ significance level က 0.026 (0.1 ထက် ငယ်) ဖြစ်နေတဲ့အတွက် variances ညီတယ်ဆိုတဲ့ hypothesis ကို ပယ်ရပါမယ်။
- ဒါကြောင့် variances မညီဘူးလို့ ယူဆပြီး alternative hypothesis ကို လက်ခံရမှာဖြစ်တဲ့အတွက် gas နဲ့ solid fuels သုံးတဲ့ အိမ်ထောင်စုတွေရဲ့ ပျမ်းမျှ heating အသုံးစရိတ်ဟာ မတူညီကြပါဘူး။

Example 4: Nonparametric Tests – Distribution အတွက် Hypothesis စမ်းသပ်ခြင်း

- **ပြဿနာ - Heating အသုံးစရိတ်ဟာ normal distribution ဖြစ်၊ မဖြစ် စမ်းသပ်ပါ။ (alpha = 5%)**
- parametric tests အများစုက normal distribution ဖြစ်တဲ့ variables တွေကို လိုအပ်တဲ့အတွက် ဒါကို အရင်စစ်ဖို့ လိုပါတယ်။
- ဒါကို စစ်ဆေးဖို့ SPSS ရဲ့ **Analyze / Nonparametric tests / 1-Sample K-S Test** ကို သုံးနိုင်ပါတယ်။

- **One-Sample Kolmogorov – Smirnov Test** procedure ဟာ variable တစ်ခုရဲ့ တွေ့ရှိရတဲ့ distribution နဲ့ သတ်မှတ်ထားတဲ့ theoretical distribution (ဥပမာ - normal distribution) တို့ကို နှိုင်းယှဉ်ပေးတာပါ။

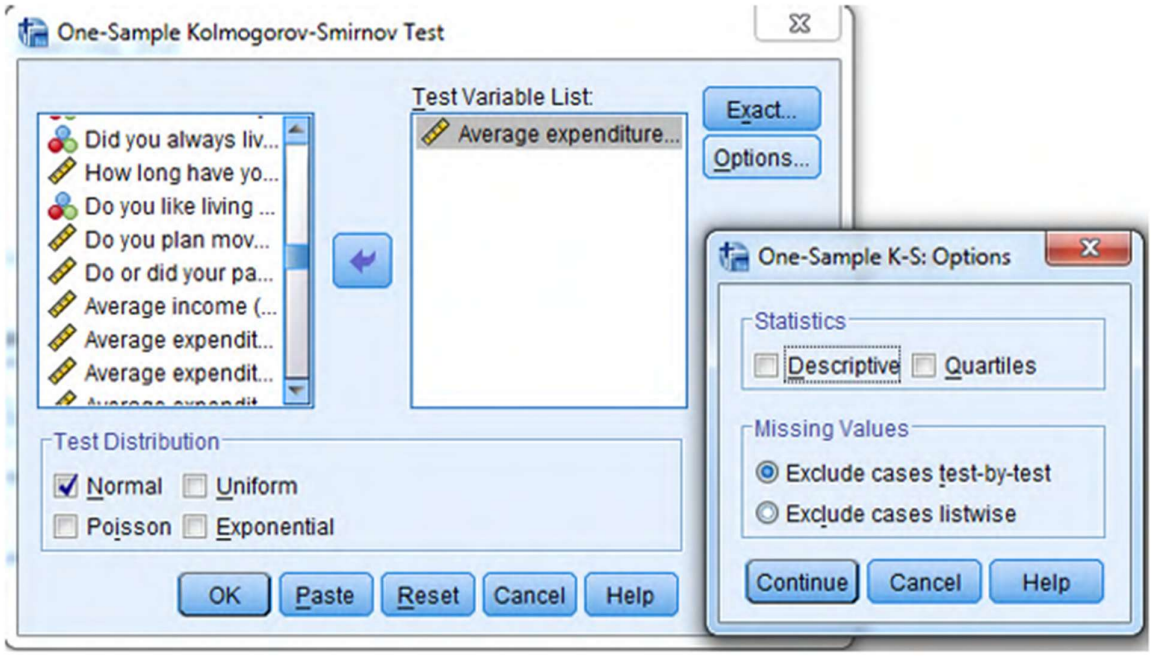


Figure 31 – One-Sample Kolmogorov–Smirnov Test Dialog Box

One-Sample Kolmogorov-Smirnov Test

		Average expenditure on heating (HUF/month)
N		193
Normal Parameters[a,b]	Mean	31727.10
	Std. Deviation	25014.693
Most Extreme Differences	Absolute	.165
	Positive	.165
	Negative	-.122
Kolmogorov-Smirnov Z		2.290
Asymp. Sig. (2-tailed)		**.000**

a. Test distribution is Normal.
b. Calculated from data.

- ရလဒ်မှာ p-value (asymp. sig.) ဟာ 0.05 ထက် နည်းနေရင် normality assumption ကို ပယ်ရမှာဖြစ်ပြီး၊ အခုရလဒ်မှာ 0.000 ဖြစ်နေတဲ့အတွက် ပျမ်းမျှ လစဉ် heating expenditure ဟာ normal distribution မဖြစ်ပါဘူး။

- နောက်တစ်နည်းအနေနဲ့ **histogram** ဆွဲပြီးတော့လည်း normal distribution ဟုတ်၊ မဟုတ် စစ်လို့ရပါတယ်။ ပုံမှာ bell-shaped curve မထွက်လာဘဲ စောင်းနေရင် normal distribution မဟုတ်ပါဘူး။

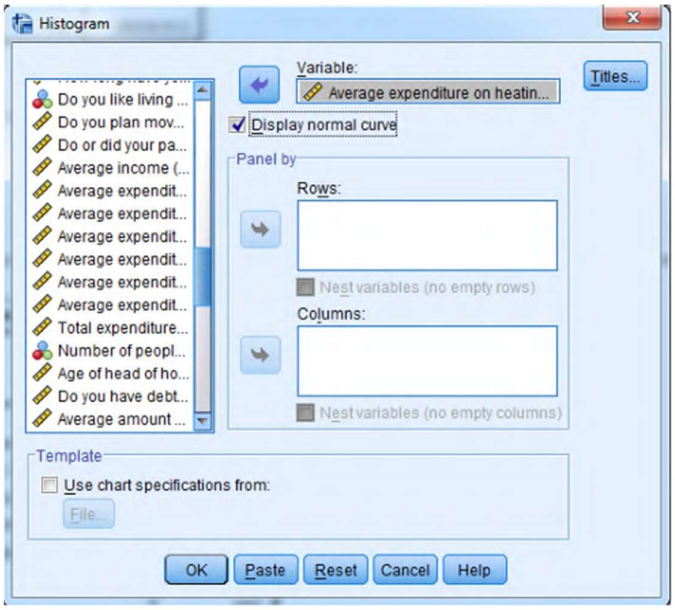


Figure 32 – Histogram Dialog Box

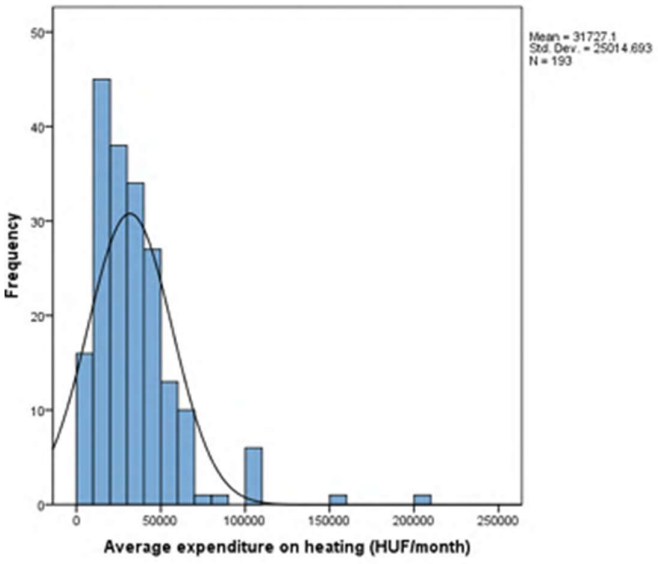


Figure 33 – Histogram

Example 5: Heating အသုံးစရိတ်ဟာ normal distribution မဖြစ်ဘူးဆိုတာ သိထားရင်၊ gas နဲ့ solid fuel သုံးတဲ့ အိမ်ထောင်စုတွေရဲ့ heating အသုံးစရိတ် ညီ၊ မညီကို စမ်းသပ်ကြည့်ရအောင်။

- variable က normal distribution မဟုတ်ရင် hypothesis စမ်းသပ်ဖို့ nonparametric test ကို သုံးရပါမယ်။
- menu ထဲကနေ **Analyze / Nonparametric tests / 2-Independent Samples** ကို ရွေးပါ။

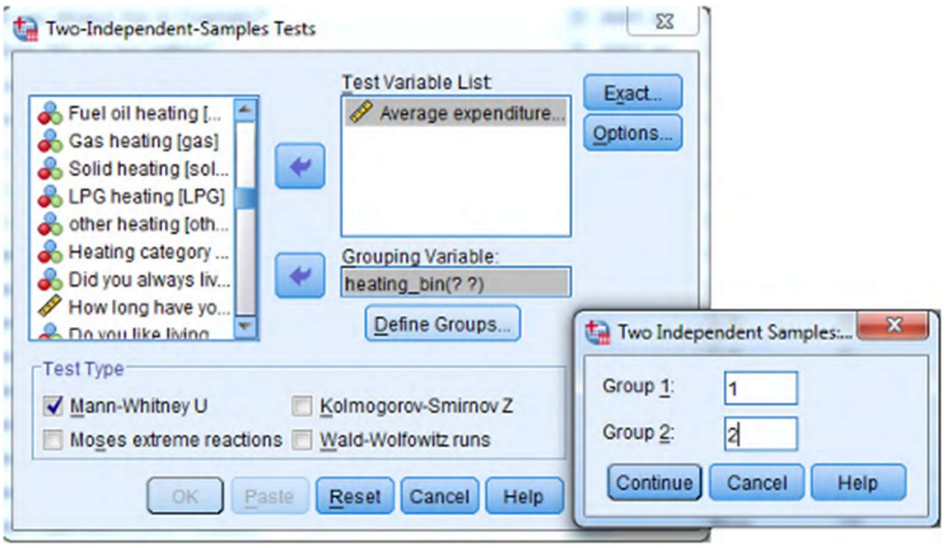


Figure 34 – Two-Independent-Samples Tests Dialog Box

- test variable နေရာမှာ heating expenditure ကို ထည့်ပြီး grouping variable မှာ heating_bin ကို ထည့်ပါ။
- **Mann-Whitney U Test** ကတော့ လူကြိုက်အများဆုံး nonparametric test ပါ။
- ရရှိလာတဲ့ p-value အရ clericals နဲ့ managers တွေရဲ့ လစာဟာ မတူညီဘူးဆိုတာကို တွေ့ရပါတယ်။ (မှတ်ချက် - manual စာအုပ်ပါ ရှင်းပြချက်မှာ လစာအကြောင်း ဖော်ပြထားသော်လည်း ဇယားထဲမှာတော့ heating အသုံးစရိတ်ကို တွက်ပြထားတာ ဖြစ်ပါတယ်။)

Test Statistics[a]

	Average expenditure on heating (HUF/month)
Mann-Whitney U	4020.000
Wilcoxon W	6576.000
Z	-.743
Asymp. Sig. (2-tailed)	.458

a. Grouping Variable: Gas or solid fuel

Chapter 6- STATISTICAL DEPENDENCE (ကိန်းဂဏာန်းဆိုင်ရာ ဆက်နွယ်မှုများ)

Example 1: အပူပေးစနစ် (heating_bin) နဲ့ household clusters တွေကြားက crosstabs ကို တည်ဆောက်ကြည့်ရအောင်။

- Crosstabs ဆိုတာကတော့ variable တစ်ခုထက်ပိုတဲ့ အချက်အလက်တွေရဲ့ ပြန့်နှံ့မှုကို contingency table (ဇယား) အနေနဲ့ ထုတ်ပေးတာပဲ ဖြစ်ပါတယ်။
- ဒီလိုလုပ်ဖို့အတွက် menu ထဲက **Analyze / Descriptive Statistics / Crosstabs** ကို ရွေးပေးပါ။

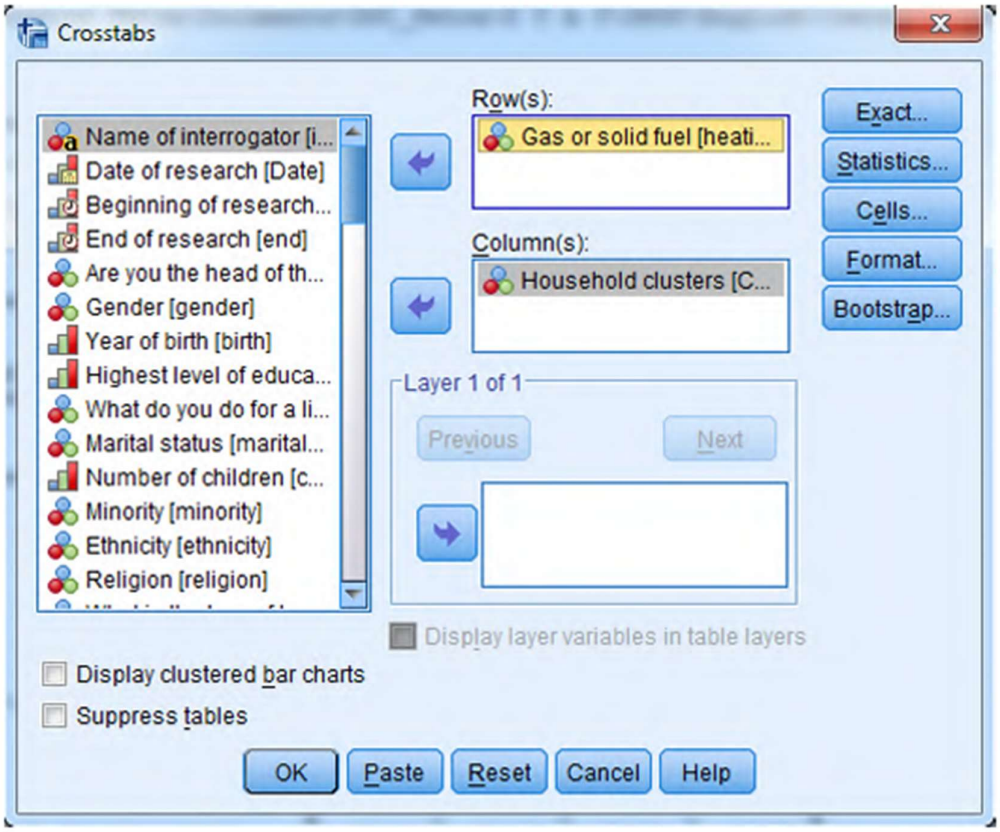


Figure 35 – Crosstabs Dialog Box

- ပြီးရင် Row နေရာနဲ့ Column နေရာတွေမှာ မိမိကြည့်ချင်တဲ့ variable တွေကို စိတ်ကြိုက်ထည့်လို့ ရပါတယ်။
- Cells ကို နှိပ်လိုက်ရင်တော့ ဇယားကွက်ထဲမှာ counts တွေအပြင် percentages တွေ ဒါမှမဟုတ် residuals တွေကိုပါ ဖော်ပြခိုင်းလို့ ရပါတယ်။

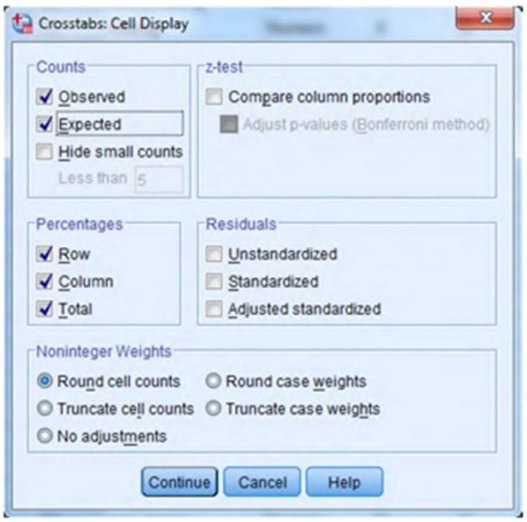


Figure 36 – Cell Display Dialog Box

- ဇယားကွက်တစ်ခုချင်းစီမှာ လက်တွေ့တွေ့ရှိရတဲ့ counts (Observed counts - fij) နဲ့ အကယ်၍ variables တွေက တစ်ခုနဲ့တစ်ခု လွတ်လပ်နေမယ် (independent ဖြစ်နေမယ်) ဆိုရင် ရှိနေရမယ့် counts (Expected counts - f*ij) တွေကို တွေ့နိုင်ပါတယ်။
- Expected frequencies ကို တွက်ချက်တဲ့အခါမှာတော့ marginal frequencies တွေကို အခြေခံပြီး တွက်ချက်တာ ဖြစ်ပါတယ်။

Gas or solid fuel * Household clusters Crosstabulation

			Household clusters			Total
			"poor"	"retired"	"well off"	
Gas or solid fuel	mainly gas	Count	6	53	12	71
		Expected Count	17.8	45.2	8.1	71.0
		% within Gas or solid fuel	8.5%	74.6%	16.9%	100.0%
		% within Household clusters	10.9%	37.9%	48.0%	32.3%
		% of Total	2.7%	24.1%	5.5%	32.3%
	solid fuel	Count	49	87	13	149
		Expected Count	37.3	94.8	16.9	149.0
		% within Gas or solid fuel	32.9%	58.4%	8.7%	100.0%
		% within Household clusters	89.1%	62.1%	52.0%	67.7%
		% of Total	22.3%	39.5%	5.9%	67.7%
Total		Count	55	140	25	220
		Expected Count	55.0	140.0	25.0	220.0
		% within Gas or solid fuel	25.0%	63.6%	11.4%	100.0%
		% within Household clusters	100.0%	100.0%	100.0%	100.0%
		% of Total	25.0%	63.6%	11.4%	100.0%

Example 2: အပူပေးစနစ် (heating_bin) နဲ့ household clusters ကြားမှာ ဆက်စပ်မှု (dependence) ရှိ၊ မရှိ စမ်းသပ်ကြည့်ရအောင်။

- ဒီနေရာမှာ variable နှစ်ခုလုံးက qualitative variables တွေ ဖြစ်နေတဲ့အတွက် ဆက်စပ်မှု ဘယ်လောက်အားကောင်းလဲဆိုတာ သိဖို့ association တိုင်းတာချက်တွေကို တွက်ရမှာပါ။
- Example 1 မှာလိုပဲ **Analyze / Descriptive Statistics / Crosstabs** ကို သွားပြီး Statistics ကို နှိပ်ပါ။

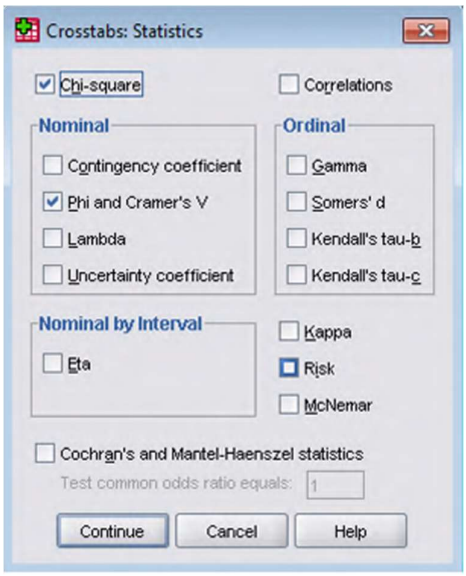


Figure 37 – Crosstabs Statistics Dialog Box

- Pearson chi-square နဲ့ likelihood-ratio တွေကို တွက်ချက်ဖို့ Chi-square ကို ရွေးပေးရပါမယ်။
- Chi-square test ဆိုတာက အမြဲတမ်း right tail test အမျိုးအစား ဖြစ်ပါတယ်။
- Pearson chi-square statistic ကို variables တွေက တစ်ခုနဲ့တစ်ခု လွတ်လပ်မှု ရှိ၊ မရှိ (independent ဖြစ်၊ မဖြစ်) ဆိုတဲ့ hypothesis ကို စမ်းသပ်ဖို့ သုံးပါတယ်။
- ဒီစမ်းသပ်မှုမှာ H0 ကတော့ variables တွေက တစ်ခုနဲ့တစ်ခု independent ဖြစ်တယ်လို့ ယူဆပြီး၊ H1 ကတော့ independent မဖြစ်ဘူးလို့ ယူဆတာပါ။
- Degrees of freedom ကတော့ row variable နဲ့ column variable တို့ရဲ့ degrees of freedom တွေ မြောက်ထားတဲ့ ရလဒ်ပဲ ဖြစ်ပါတယ်။

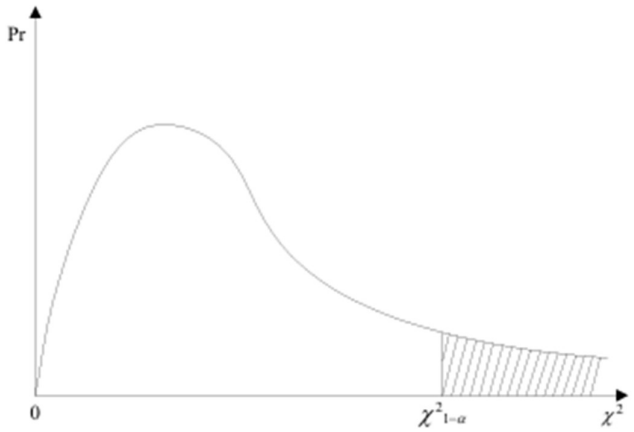


Figure 38 – Chi-square Distribution

- ယေဘုယျအားဖြင့် significance value က 0.05 ထက် နည်းနေမယ်ဆိုရင် ဆက်စပ်မှု ရှိတယ် (significant ဖြစ်တယ်) လို့ သတ်မှတ်ပါတယ်။
- Nominal data တွေအတွက်ဆိုရင် Cramer's V ကို သုံးနိုင်ပြီး သူက chi-square ပေါ်မှာ အခြေခံထားတဲ့ association တိုင်းတာချက် ဖြစ်ပါတယ်။
- Cramer's V တန်ဖိုးက 0.3 ထက် နည်းရင် dependence အားနည်း (weak) တယ်၊ 0.3 နဲ့ 0.7 ကြားဆိုရင် အလယ်အလတ် (medium-strong) ရှိပြီး၊ 0.7 ထက် များရင်တော့ dependence အားကောင်း (strong) တယ်လို့ သတ်မှတ်ပါတယ်။

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	16.311[a]	2	.000
Likelihood Ratio	18.454	2	.000
Linear-by-Linear Association	14.754	1	.000
N of Valid Cases	220		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 8.07.

- ရရှိလာတဲ့ chi-square results အရ variables တွေက independent မဖြစ်ပါဘူး။

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.272	.000
	Cramer's V	.272	.000
N of Valid Cases		220	

- အပူပေးစနစ်နဲ့ household clusters ကြားမှာ အားနည်းတဲ့ ဆက်စပ်မှု (weak dependence) ရှိနေတယ်ဆိုတာကို significance level တိုင်းမှာ လက်ခံနိုင်ပါတယ်။

Example 3: အပူပေးစနစ် (heating_bin) နဲ့ ပျမ်းမျှ အပူပေးစရိတ် (average expenditures on heating) ကြားမှာ ဆက်စပ်မှု ရှိ၊ မရှိ ကြည့်ရအောင်။

- ဒီနေရာမှာတော့ dependent variable က quantitative ဖြစ်ပြီး independent variable က categorical ဖြစ်နေတဲ့အတွက် ဒါကို mixed dependence လို့ ခေါ်ပါတယ်။
- ဒါကို သိနိုင်ဖို့ menu ထဲက **Analyze / Compare Means / Means** ကို သွားရပါမယ်။

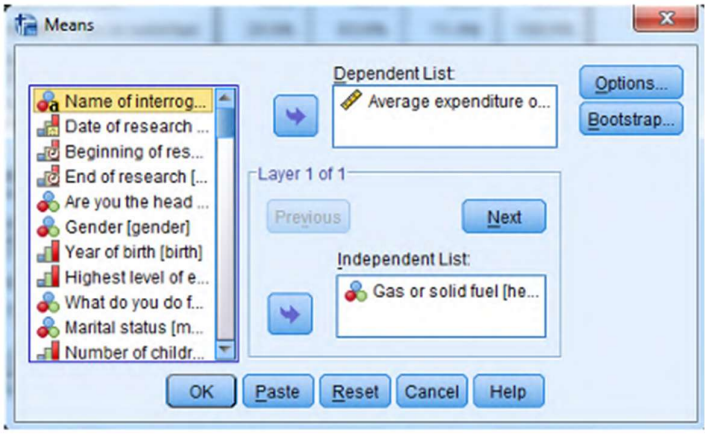


Figure 39 – Means Dialog Box

- Options ကို နှိပ်ပြီး subgroups တွေရဲ့ means နဲ့ standard deviations တွေကို ကြည့်လို့ရသလို mixed dependence တိုင်းတာချက်တွေကိုလည်း ဒီမှာပဲ ရွေးလို့ရပါတယ်။

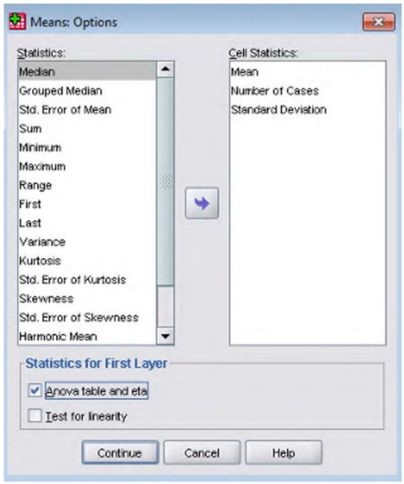


Figure 40 – Means Options Dialog Box

Report

Average expenditure on heating (HUF/month)

Gas or solid fuel	Mean	N	Std. Deviation
mainly gas	28300.00	71	16660.861
solid fuel	33917.60	121	28697.218
Total	31840.26	192	25030.512

- ANOVA table ထဲက F-statistic ကို ကြည့်ပြီး ဆက်စပ်မှုက တကယ်ပဲ ရှိတာလားဆိုတာ ဆုံးဖြတ်ပါတယ်။
- F-value ကြီးပြီး significance level (Sig.) သေးနေမယ်ဆိုရင် ရလဒ်က တိုက်ဆိုင်မှုကြောင့် မဟုတ်ဘဲ variables တွေကြားမှာ ခိုင်မာတဲ့ ဆက်စပ်မှု ရှိနေတာကို ပြပါတယ်။ ပုံမှန်အားဖြင့် Sig. က 0.05 ထက် ငယ်ရပါမယ်။

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
Average expenditure on heating (HUF/month) * Gas or solid fuel	Between Groups	(Combined)	1.412E9	1	1.412E9	2.26	.134
	Within Groups		1.183E11	190	6.224E8		
	Total		1.197E11	191			

- အခု ရလဒ်အရတော့ အပူပေးစနစ်နဲ့ အပူပေးစရိတ်ကြားမှာ သိသာထင်ရှားတဲ့ ဆက်စပ်မှု (significant relationship) မရှိတာကို တွေ့ရပါတယ်။

Measures of Association

	Eta	Eta Squared
Average expenditure on heating (HUF/month) * Gas or solid fuel	.109	.012

- Eta (H-measure) တန်ဖိုးအရvariables တွေကြား ဆက်စပ်မှုက အားနည်း (weak) ပါတယ်။
- Eta squared (H2-measure) တန်ဖိုးအရ အပူပေးစရိတ်ရဲ့ ကွဲပြားမှု (variance) ၁.၂ ရာခိုင်နှုန်းကိုပဲ အပူပေးစနစ် အမျိုးအစားနဲ့ ရှင်းပြနိုင်တာ တွေ့ရပါတယ်။

Chapter 7. CORRELATION AND LINEAR REGRESSION (ဆက်စပ်မှုနှင့် မျဉ်းဖြောင့်ပုံစံခန့်မှန်းခြင်း)

Example 1: စုစုပေါင်းအသုံးစရိတ်နဲ့ ပျမ်းမျှဝင်ငွေကြားမှာ ဆက်စပ်မှု ရှိ၊ မရှိ ကြည့်ရအောင်။

- အစပြုလစာနဲ့ လက်ရှိလစာ (ဥပမာထဲမှာတော့ အသုံးစရိတ်နဲ့ ဝင်ငွေ) တို့ဟာ quantitative variables တွေ ဖြစ်တဲ့အတွက် သူတို့ကြားက ဆက်စပ်မှုကို **correlation** လို့ ခေါ်ပါတယ်၊ ဒါကြောင့် correlation coefficient ကို တွက်ချက်ရမှာပါ။
- ဒီလိုလုပ်ဖို့ Menu ထဲက **Analyze / Correlate / Bivariate** ကို ရွေးပေးပါ။
- Quantitative ဖြစ်ပြီး normal distribution ရှိတဲ့ variable တွေအတွက်ဆိုရင် **Pearson correlation coefficient** ကို ရွေးချယ်ပေးရမှာပါ။
- တကယ်လို့ data တွေက normal distribution မဟုတ်ရင် ဒါမှမဟုတ် အဆင့်လိုက် ခွဲခြားထားတဲ့ categories တွေ ဖြစ်နေရင်တော့ rank orders တွေကို တိုင်းတာပေးတဲ့ **Kendall's tau-b** သို့မဟုတ် **Spearman** ကို ရွေးချယ်ရပါမယ်။

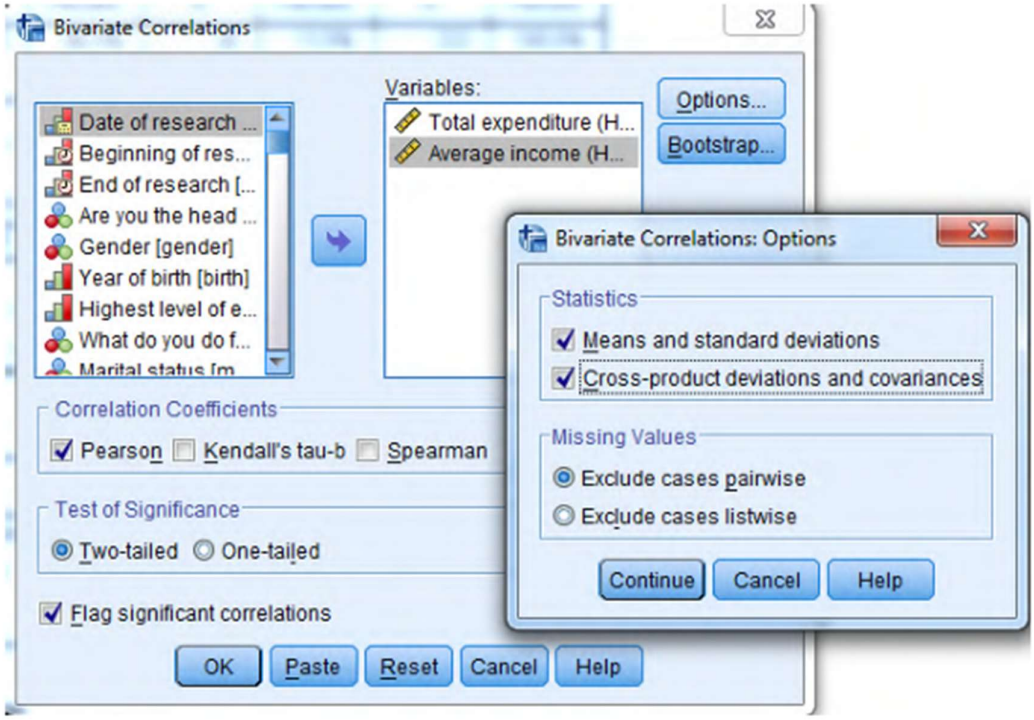


Figure 41 – Bivariate Correlations Dialog Box

- လိုအပ်ရင် **Options** ကို နှိပ်ပြီး **covariance** ကိုလည်း ဖော်ပြခိုင်းလို့ ရပါတယ်။ Covariance ဆိုတာကတော့ variable နှစ်ခုကြား ဆက်စပ်မှုကို standardize မလုပ်ဘဲ

တိုင်းတာတာဖြစ်ပြီး ဆက်စပ်မှုရဲ့ လားရာ (positive ဒါမှမဟုတ် negative relationship) ကို ပြပေးတာပါ။

		Total expenditure (HUF/month)	Average income (HUF/month)
Total expenditure (HUF/month)	***Pearson Correlation***	1	**0.539(**)**
	Sig. (2-tailed)		**0.000**
	Covariance	3254171744.937	**2147784135.965**
	N	222	217
Average income (HUF/month)	Pearson Correlation	0.539(**)	1
	Sig. (2-tailed)	0.000	
	Covariance	2147784135.965	4814606330.816
	N	217	217

- ရရှိလာတဲ့ ရလဒ်တွေအရ (r = 0.539, Sig. = 0.000) အသုံးစရိတ်နဲ့ ဝင်ငွေကြားမှာ သိသာထင်ရှားပြီး အပေါင်းလက္ခဏာဆောင်တဲ့ အလယ်အလတ်စား ဆက်စပ်မှု (positive, medium-strong and significant relationship) ရှိနေတာကို တွေ့ရပါတယ်။

Example 2: ပျမ်းမျှ လစဉ်ဝင်ငွေနဲ့ စုစုပေါင်းအသုံးစရိတ်ကြားက linear relation ကို ရှာကြည့်ရအောင်။

- **Linear regression** ဆိုတာကတော့ independent variable တစ်ခုကို သုံးပြီး dependent variable ရဲ့ တန်ဖိုးကို အကောင်းဆုံး ခန့်မှန်းပေးနိုင်မယ့် linear equation တစ်ခုရဲ့ coefficients တွေကို တွက်ချက်ပေးတာပါ။
- အရင်ဆုံး ဘယ် variable က တခြားတစ်ခုအပေါ် လွှမ်းမိုးမှု ရှိသလဲဆိုတာ စဉ်းစားရပါမယ်၊ ဒီနေရာမှာ ဝင်ငွေက အသုံးစရိတ်အပေါ် လွှမ်းမိုးမှု ရှိနိုင်တဲ့အတွက် **Dependent variable** က total expenditure ဖြစ်ပြီး **Independent variable** ကတော့ average monthly income ဖြစ်ပါတယ်။
- တွက်ချက်ဖို့အတွက် Menu ထဲက **Analyze / Regression / Linear** ကို သွားပါ။

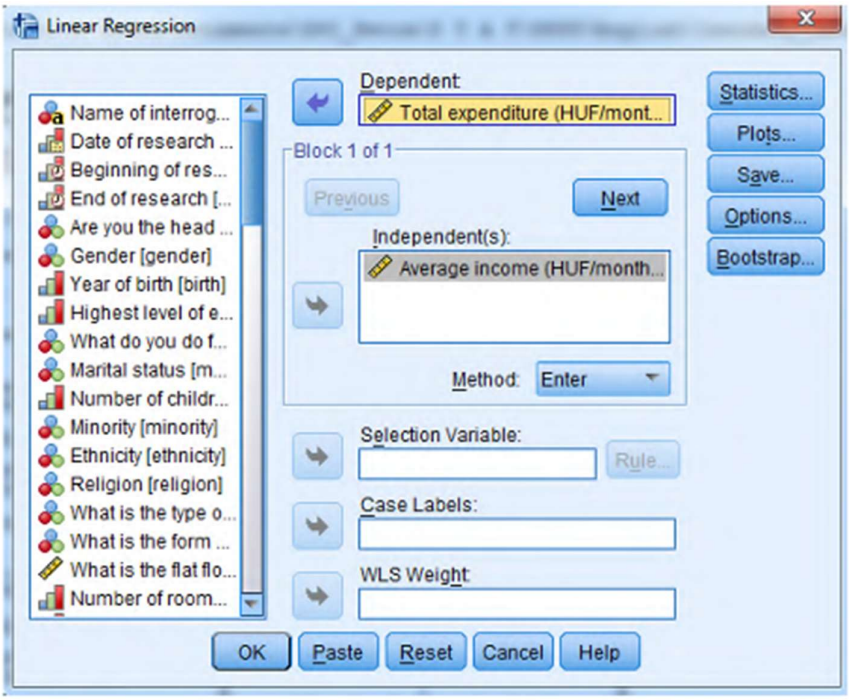


Figure 42 – Linear Regression Dialog Box

- လိုအပ်ရင် **Statistics** ကို နှိပ်ပြီး **covariance matrix** နဲ့ correlation coefficients ဇယားတွေကို ထပ်မံ ဖော်ပြခိုင်းလို့ ရပါတယ်။

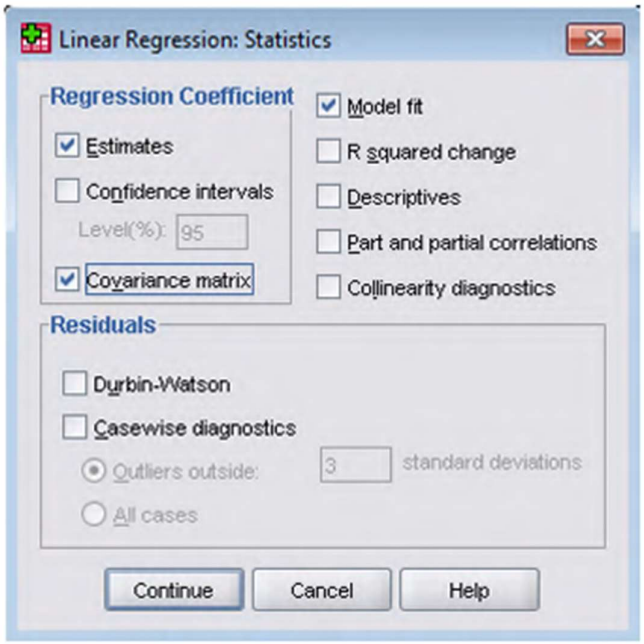


Figure 43 – Linear Regression: Statistics Dialog Box

- **ANOVA table** ကိုတော့ regression model တစ်ခုလုံး တကယ်ပဲ အသုံးဝင်၊ မဝင် (overall regression significance) ကို စစ်ဆေးဖို့ သုံးပါတယ်။ Significance level က သုညနီးပါး ဖြစ်နေရင် (0.05 ထက် ငယ်ရင်) model က ခိုင်မာတယ်လို့ ပြောနိုင်ပါတယ်။

ANOVA table for bivariate regression model

Model	Sum of Squares	df	Mean Squares	F
Regression	$S_{\hat{y}} = \Sigma(\hat{y}_i - \bar{y})^2$	1	$S_{\hat{y}}$	$F = \frac{S_{\hat{y}}}{S_e/(n-2)}$
Residual	$S_e = \Sigma(y_i - \hat{y})^2$	n-2	$s_e^2 = S_e/(n-2)$	
Total	$S_y = \Sigma(y_i - \bar{y})^2$	n-1		

ANOVA [b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2.070E11	1	2.070E11	88.012	**.000**[a]
	Residual	5.056E11	215	2.351E9		
	Total	7.125E11	216			

a. Predictors: (Constant), Average income (HUF/month)

b. Dependent Variable: Total expenditure (HUF/month)

- Linear regression ရဲ့ coefficients တွေကို အခုလို အဓိပ္ပာယ်ဖွင့်နိုင်ပါတယ်-

Coefficients [a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	**47355.059**	6534.580		7.247	**0.000**
	Average income (HUF/month)	**0.446**	.048	.539	9.381	**0.000**

  - **Constant:** ဒါကတော့ x=0 ဖြစ်တဲ့အခါ ကျရောက်တဲ့ intercept အမှတ်ပါ။ ယေဘုယျအားဖြင့်တော့ သူက အဓိပ္ပာယ် သိပ်မရှိတဲ့အတွက် ထည့်ပြီး အဓိပ္ပာယ်မဖွင့်ပါဘူး။ ဒါပေမဲ့ ဒီဥပမာအရဆိုရင်တော့ ဝင်ငွေ လုံးဝမရှိတဲ့အခါ ပျမ်းမျှအသုံးစရိတ်က 47,355.059 HUF ရှိနေမယ်လို့ ဆိုလိုတာပါ။
  - **b1 coefficient:** Independent variable တစ်ခု တိုးလာရင် dependent variable ဘယ်လောက်အထိ တိုးလာမလဲ (ဒါမှမဟုတ် လျော့သွားမလဲ) ဆိုတာကို ပြောတာပါ။

ဒီမှာဆိုရင် ဝင်ငွေ ၁ HUF တိုးလာတိုင်း အသုံးစရိတ်က 0.446 HUF တိုးလာမယ်လို့ ခန့်မှန်းနိုင်ပါတယ်။

- **t-test:** model ထဲက parameters တွေ တကယ်ပဲ အသုံးဝင်ရဲ့လားဆိုတာကို စစ်ဆေးတာပါ။ p-value က 0.05 ထက် နည်းရင် null hypothesis ကို ပယ်ပြီး significant ဖြစ်တယ်လို့ သတ်မှတ်ပါတယ်။

Example 3: b1 parameter အတွက် 95% confidence interval ကို ရှာကြည့်ရအောင်။

- **Analyze / Regression / Linear / Statistics** ထဲကို သွားပါ။
- b coefficients တွေရဲ့ confidence interval ကို သိချင်ရင် **Confidence intervals** ဆိုတဲ့ box လေးကို အမှန်ခြစ်ပေးပြီး လိုအပ်တဲ့ confidence level (ဥပမာ- 95%) ကို သတ်မှတ်ပေးပါ။

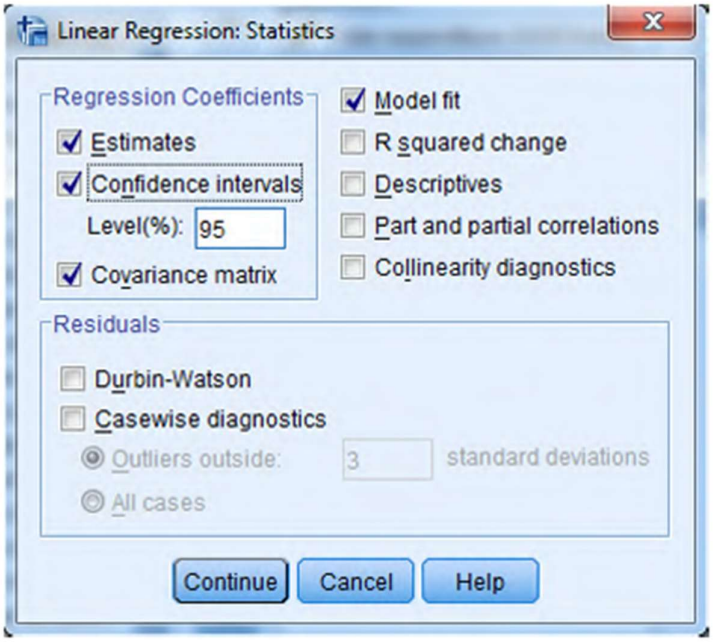


Figure 44 – Linear Regression Dialog Box: Statistics for Estimating the Coefficients

- ရရှိလာတဲ့ ရလဒ်အရ b1 coefficient ဟာ 95% confidence level မှာ 0.352 နဲ့ 0.540 ကြားမှာ ရှိနေပါတယ်။

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	47355.059	6534.580		7.247	.000	34475.016	60235.101
	Average income (HUF/month)	.446	.048	.539	9.381	.000	.352	.540

Example 4: Dependent variable (total expenditure) အတွက် 95% confidence interval ကို ရှာကြည့်ရအောင်။

- **Analyze / Regression / Linear / Save** ထဲကို သွားပါ။

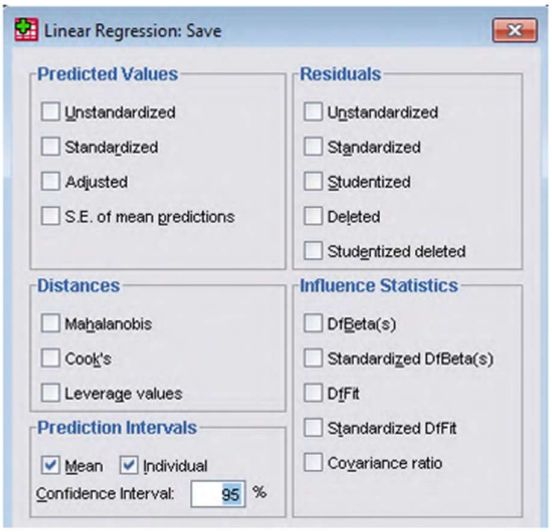


Figure 45 – Linear Regression: Save for Prediction Intervals

- **Prediction Intervals** အောက်မှာရှိတဲ့ Mean ဒါမှမဟုတ် Individual ဆိုတာကို ရွေးပြီး 95% confidence interval လို့ ရိုက်ထည့်ပါ။

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	**54 938.72**	**225 794.08**	100 312.41	30 953.53	217
Std. Predicted Value	-1.466	4.054	.000	1.000	217
Standard Error of Predicted Value	3292.386	13 774.578	4378.560	1584.899	217
Adjusted Predicted Value	53 977.68	238 686.89	100 414.08	31 347.65	217

a. Dependent Variable: Total expenditure (HUF/month)

- ရလဒ်အရ 95% confidence level မှာ ဝန်ထမ်းတွေရဲ့ (အိမ်ထောင်စုတွေရဲ့) အသုံးစရိတ်ဟာ 54,938.72 နဲ့ 225,794.08 HUF ကြားမှာ ရှိနေမှာ ဖြစ်ပါတယ်။

Chapter 8. MULTIPLE CORRELATION AND LINEAR REGRESSION

(ဆက်စပ်မှုအမြောက်အများနှင့် မျဉ်းဖြောင့်ပုံစံ ခန့်မှန်းခြင်း)

Example 1: ပျမ်းမျှဝင်ငွေ၊ စုစုပေါင်းအသုံးစရိတ်နဲ့ အိမ်ထောင်စုဝင် ဦးရေကြားမှာ ဆက်စပ်မှု ရှိ၊ မရှိ ကြည့်ရအောင်။

- ဒီလို ဆက်စပ်မှုတွေကို ကြည့်ဖို့အတွက် Menu ထဲက **Analyze / Correlate / Bivariate** ကို သွားပြီး **Pearson Correlation Coefficient** ကို ရွေးပေးရပါမယ်။
- Variables တွေအနေနဲ့ average income, total expenditure နဲ့ number of people living in the household တို့ကို ထည့်သွင်းရမှာ ဖြစ်ပါတယ်။

Correlations

		Average income	Total expenditure	Number of people living in the household
Average income	Pearson Correlation	1	**0.539****	**0.293****
	Sig. (2-tailed)		0.000	0.034
Total expenditure	Pearson Correlation	0.539**	1	**0.373****
	Sig. (2-tailed)	0.000		0.000
Number of people living in the household	Pearson Correlation	.293**	0.373**	1
	Sig. (2-tailed)	0.034	0.000	

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

- ရရှိလာတဲ့ ရလဒ်တွေ (Correlations table) ကို အခုလို အဓိပ္ပာယ်ဖွင့်နိုင်ပါတယ် -
    - ပျမ်းမျှဝင်ငွေနဲ့ စုစုပေါင်းအသုံးစရိတ်ကြားက ဆက်စပ်မှု (r = 0.539) ကတော့ သိသာထင်ရှားပြီး အပေါင်းလက္ခဏာဆောင်တဲ့ အလယ်အလတ်စား ဆက်စပ်မှု (Positive, medium-strong, significant relationship) ဖြစ်ပါတယ်။
    - ပျမ်းမျှဝင်ငွေနဲ့ အိမ်ထောင်စုဝင် ဦးရေကြားက ဆက်စပ်မှု (r = 0.293) ကတော့ သိသာထင်ရှားပေမယ့် အားနည်းတဲ့ အပေါင်းလက္ခဏာဆောင် ဆက်စပ်မှု (Positive, weak, significant relationship) ဖြစ်ပါတယ်။
    - စုစုပေါင်းအသုံးစရိတ်နဲ့ အိမ်ထောင်စုဝင် ဦးရေကြားက ဆက်စပ်မှု (r = 0.373) ကတော့ သိသာထင်ရှားပြီး အလယ်အလတ်စား အပေါင်းလက္ခဏာဆောင် ဆက်စပ်မှု (Positive, medium-strong, significant relationship) ရှိနေတာကို တွေ့ရပါတယ်။

Example 2: ပျမ်းမျှဝင်ငွေ၊ စုစုပေါင်းအသုံးစရိတ်နဲ့ အိမ်ထောင်စုဝင် ဦးရေကြားက linear relation ကို ရှာကြည့်ရအောင်။

- ဒါကို တွက်ချက်ဖို့ Menu ထဲက **Analyze / Regression / Linear** ကို သွားပါ။
- **Dependent variable** နေရာမှာ Total expenditure ကို ထည့်ပြီး၊ **Independents variable** နေရာမှာတော့ Average income နဲ့ Number of people living in the household တို့ကို ထည့်ပေးရပါမယ်။

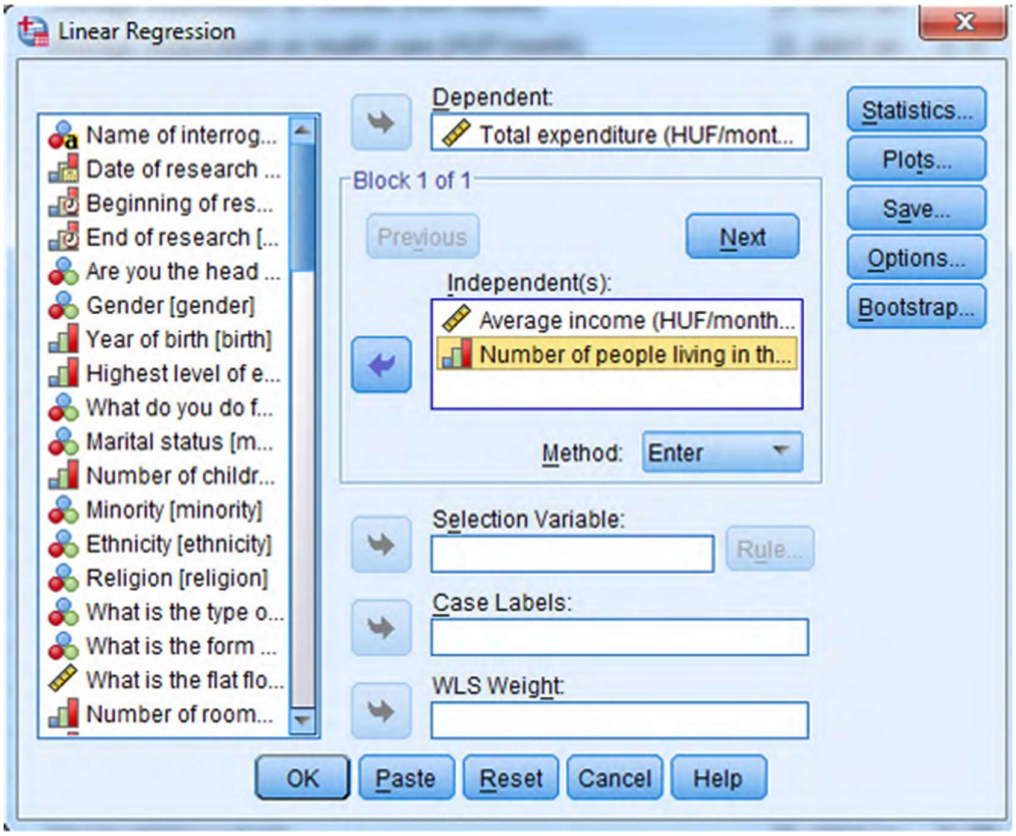


Figure 46 – Multiple Linear Regression

- Multiple linear regression မှာဆိုရင် လိုက်နာရမယ့် အခြေခံယူဆချက် (conditions) တွေ ပိုများလာပါတယ် -
    - **error term** တွေအတွက် Assumptions တွေကတော့ - error တွေဟာ normal distribution ဖြစ်ရမယ်၊ error ရဲ့ expected value က သုည ဖြစ်ရမယ် (E(ε)=0)။ variance တွေက တူညီရမယ် (Homoscedasticity)။ ပြီးတော့ တစ်ခုနဲ့တစ်ခု ဆက်စပ်မှု မရှိရပါဘူး (no autocorrelation)။

- **Independent Variables** တွေအတွက် Assumptions တွေကတော့ - independent variable တွေဟာ random မဖြစ်ရဘူး၊ ပြီးတော့ variable တွေကြားမှာ တစ်ခုနဲ့တစ်ခု ဆက်စပ်မှု မရှိရပါဘူး (No multicollinearity)။

- ဒီယူဆချက်တွေကို SPSS မှာ စစ်ဆေးလို့ ရပေမယ့်၊ ဒါက MSc တန်းတွေမှာ သင်ရမယ့် အကြောင်းအရာ ဖြစ်တဲ့အတွက် ဒီ tutorial မှာတော့ အသေးစိတ် မပြတော့ပါဘူး။

- ရရှိလာတဲ့ Linear regression ရလဒ်တွေကို အခုလို ဆန်းစစ်နိုင်ပါတယ် -

Model Summary [b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	**0.581**[a]	**0.338**	**0.332**	46 948.77	1.780

a. Predictors: (Constant), Average income (HUF/month), Number of people living in the household

b. Dependent Variable: Total expenditure (HUF/month)

  - **R (coefficient of correlation):** အမှီပြောင်းလဲကိန်း (total expenditure) နဲ့ အမှီအခိုကင်း ပြောင်းလဲကိန်းတွေ (average income & number of people) ကြားမှာ အလယ်အလတ်စား ဆက်စပ်မှု ရှိပါတယ်။

  - **R2 (coefficient of determination):** အသုံးစရိတ်ရဲ့ ကွဲပြားမှု (variance) ၃၃.၈ ရာခိုင်နှုန်းကို ဝင်ငွေနဲ့ လူဦးရေတို့က ရှင်းပြနိုင်ပါတယ်။

  - **Adjusted R2:** sample size နဲ့ parameter အရေအတွက်အလိုက် ပြုပြင်တွက်ချက်ပြီးတဲ့အခါမှာတော့ အသုံးစရိတ်ရဲ့ ၃၃.၂ ရာခိုင်နှုန်းကို ရှင်းပြနိုင်တယ်လို့ ဆိုလိုတာပါ။

- **ANOVA table** အရ significance level က သုညနီးပါး ဖြစ်နေတဲ့အတွက် ဒီ multiple regression model တစ်ခုလုံးဟာ ခိုင်မာမှုရှိတယ် (significant ဖြစ်တယ်) လို့ ပြောနိုင်ပါတယ်။

ANOVA table for multiple regression model

Model	Sum of Squares	df	Mean Squares	F
Regression	$S_{\hat{y}} = \Sigma(\hat{y}_i - \bar{y})^2$	p^4	$S_{\hat{y}}$	$F = \frac{S_{\hat{y}}}{S_e/(n - p - 1)}$
Residual	$S_e = \Sigma(y_i - \hat{y})^2$	n-p-1	$s_e^2 = S_e/(n - p - 1)$	
Total	$S_y = \Sigma(y_i - \bar{y})^2$	n-1		

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2.408E11	2	1.204E11	54.627	**0.000[a]**
	Residual	4.717E11	214	2.204E9		
	Total	7.125E11	216			

a. Predictors: (Constant), Average income (HUF/month), Number of people living in the household

b. Dependent Variable: Total expenditure (HUF/month)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	32105.279	7427.351		4.323	.000	17465.143	46745.415		
	Average income (HUF/month)	.391	.048	.472	8.116	.000	.296	.486	.914	1.094
	Number of people living in the household	8334.457	2126.456	.228	3.919	.000	4142.976	12525.938	.914	1.094

a. Dependent Variable: Total expenditure (HUF/month)

- Coefficients (parameters) တွေကိုတော့ အခုလို အဓိပ္ပာယ်ဖွင့်နိုင်ပါတယ် -
    - **b0:** ဒါကိုတော့ ယေဘုယျအားဖြင့် အသေးစိတ် မဆန်းစစ်ပါဘူး (Variables တွေ အားလုံး သုညဖြစ်တဲ့အခါ ထွက်လာတဲ့ ပျမ်းမျှတန်ဖိုး ဖြစ်လို့ပါ)။
    - **b1:** တခြား variable တွေကို ပုံသေထားပြီး ပျမ်းမျှဝင်ငွေ ၁ HUF တိုးလာတိုင်း၊ စုစုပေါင်းအသုံးစရိတ်က 0.391 HUF တိုးလာဖို့ မျှော်လင့်နိုင်ပါတယ်။
    - **b2:** တခြား variable တွေကို ပုံသေထားပြီး အိမ်ထောင်စုဝင် ၁ ယောက် တိုးလာတိုင်း၊ စုစုပေါင်းအသုံးစရိတ်က 8334.457 HUF အထိ တိုးလာဖို့ မျှော်လင့်နိုင်ပါတယ်။

Chapter-9. CURVILINEAR REGRESSION (ကွေးညွှတ်သော မျဉ်းပုံစံဖြင့် ခန့်မှန်းတွက်ချက်ခြင်း)

Example (ဥပမာ)

- Total expenditure နဲ့ average income ကြားက ဆက်စပ်မှုမှာ ဘယ် regression model က အကိုက်ညီဆုံး ဖြစ်မလဲဆိုတာ ကြည့်ရအောင်။

လုပ်ဆောင်ပုံ အဆင့်ဆင့်

- regression အမျိုးအစားတွေကို curve estimation နဲ့ စမ်းသပ်ကြည့်ဖို့အတွက် menu ထဲက **Analyze / Regression / Curve Estimation** ကို သွားပါ။
- **Dependent variable** နေရာမှာ Total expenditure ကို ထည့်ပြီး၊ **Independent variable** နေရာမှာတော့ Average income ကို ထည့်ပေးပါ။
- ပြီးရင် Models တွေထဲကမှ **Linear, Power** နဲ့ **Compound** ဆိုတာတွေကို ရွေးပေးရပါမယ်။

အကောင်းဆုံး ကိုက်ညီမယ့် regression model ကို စမ်းသပ်ဖို့ နည်းလမ်း ၃ ခု ရှိပါတယ် -

၁။ **Plot model** ('Plot Model' ကို ရွေးချယ်ပါ) - သူကတော့ dependent variable တန်ဖိုးတွေနဲ့ ရွေးချယ်ထားတဲ့ model တစ်ခုချင်းစီကို independent variable နဲ့ ယှဉ်ပြီး ပုံဆွဲပြပေးမှာပါ။ ဒီနည်းနဲ့ ဘယ် model ကို သုံးသင့်လဲဆိုတာ ဆုံးဖြတ်နိုင်ပါတယ်။ တကယ်လို့ variables တွေက linear ပုံစံ ဆက်စပ်နေတယ်ဆိုရင် simple linear regression model ကို သုံးပါ။ တကယ်လို့ variables တွေက linear မဟုတ်ဘဲ ကွေးညွှတ်နေတယ်ဆိုရင်တော့ curvilinear regression ကို သုံးရပါမယ်။ အသုံးအများဆုံး curvilinear regression တွေကတော့ power နဲ့ compound regressions တွေပဲ ဖြစ်ပါတယ်။

၂။ **ANOVA table** ('Display ANOVA table' ကို ရွေးချယ်ပါ) - ရွေးချယ်ထားတဲ့ model တစ်ခုချင်းစီအတွက် summary analysis of variance ဇယားကို ပြပေးပါလိမ့်မယ်။ Significance level က 0.05 ထက် နည်းနေမယ်ဆိုရင်တော့ အဲဒီ model က ဒေတာတွေနဲ့ ကိုက်ညီမှုရှိတယ် (well-fitted ဖြစ်တယ်) လို့ ပြောနိုင်ပါတယ်။

၃။ **Model summary** - R2 (coefficient of determination) က linear regression မှာ model ဘယ်လောက်အထိ အံဝင်ခွင်ကျ ဖြစ်လဲ (goodness-of-fit) ဆိုတာကို ပြပေးပါတယ်။ R2 တန်ဖိုး ပိုများလေလေ model က ဒေတာတွေနဲ့ ပိုပြီး အံဝင်ခွင်ကျ ဖြစ်လေလေပါပဲ။

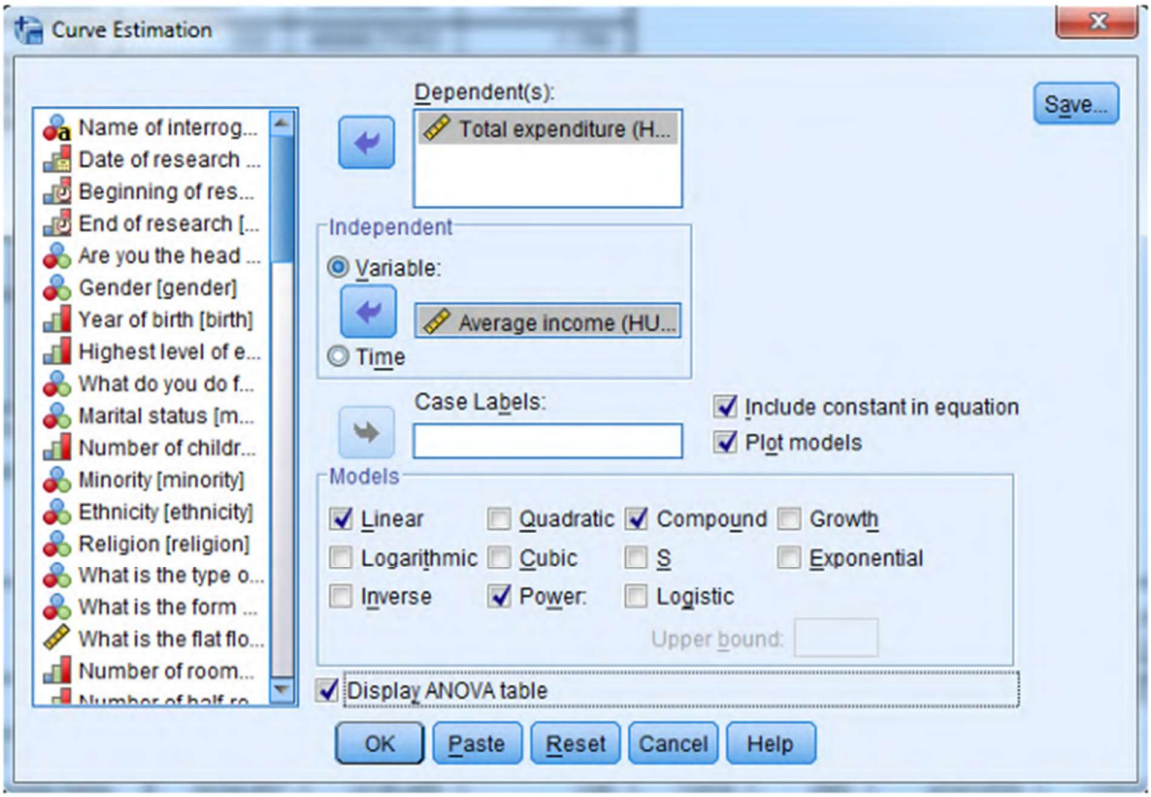


Figure 47 – Curve Estimation Dialog Box

Linear Regression Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
0.539	0.290	0.287	48 491.495

The independent variable is Average income (HUF/month).

Compound Regression Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
0.515	0.265	0.262	0.580

The independent variable is Average income (HUF/month).

Power Regression Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
0.574	**0.330**	0.326	0.554

The independent variable is Average income (HUF/month).

ရလဒ်များကို ဆန်းစစ်ခြင်း

- Output window မှာ ရလဒ်တွေ ပေါ်လာပါလိမ့်မယ်။ အခု ဥပမာမှာတော့ အခြေအနေအားလုံးအတွက် F-tests ရဲ့ significance levels တွေက 0 ဖြစ်နေပါတယ်။ ဒါကြောင့် model ဘယ်လောက် အံဝင်ခွင်ကျ ဖြစ်လဲဆိုတဲ့ goodness-of-fit ရလဒ်တွေကိုပဲ နှိုင်းယှဉ်ကြည့်ရပါမယ်။
- **Linear Regression Model Summary** အရ R2 က 0.290 ရှိပါတယ်။
- **Compound Regression Model Summary** အရ R2 က 0.265 ရှိပါတယ်။
- **Power Regression Model Summary** အရ R2 က 0.330 ရှိပါတယ်။
- ဒီ goodness-of-fit ရလဒ်တွေအရ Power regression model က ဒေတာတွေနဲ့ အကိုက်ညီဆုံး ဖြစ်နေတာကို တွေ့ရပါတယ်။
- နောက်တစ်နည်းအနေနဲ့ plot model (ပုံကားချပ်) ကိုလည်း ကြည့်နိုင်ပါတယ်။ ဒါပေမဲ့ ပုံကိုပဲ ကြည့်ပြီး ဘယ် function က အကိုက်ညီဆုံးလဲဆိုတာ ဆုံးဖြတ်ဖို့တော့ နည်းနည်း ခက်ခဲပါလိမ့်မယ်။

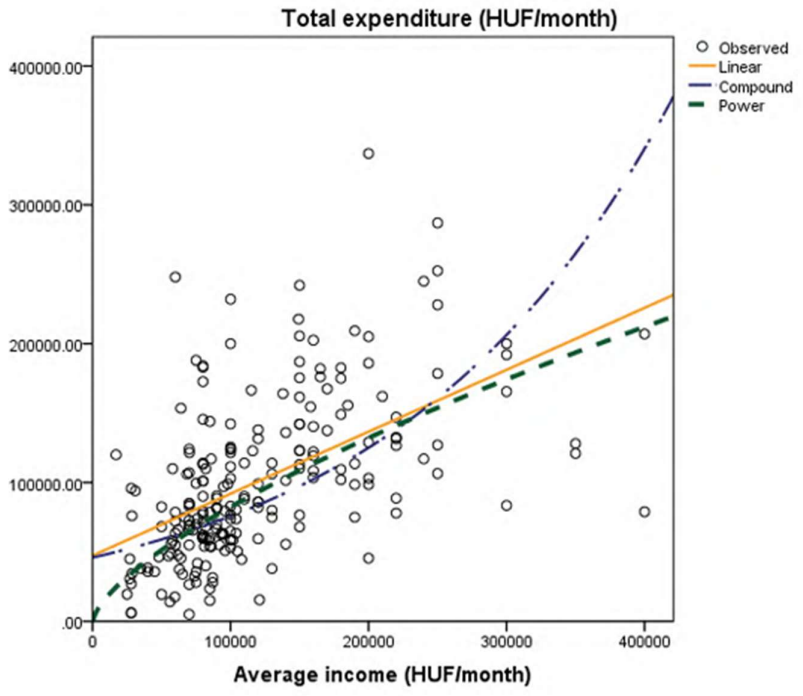


Figure 48 – Curve Fit

Parameters များကို အဓိပ္ပာယ်ဖွင့်ဆိုခြင်း

ဒေတာတွေက power regression function နဲ့ ဆင်တူတဲ့အတွက် parameter တွေကို အခုလို အဓိပ္ပာယ်ဖွင့်ကြည့်ရအောင် -

- ဒီနေရာမှာ dependent variable က **ln(Total expenditures)** ဖြစ်ပါတယ်။

Coefficients

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
ln(Average income)	**0.683**	0.066	0.574	10.279	**0.000**
(Constant)	**31.810**	24.383		1.305	**0.193**

The dependent variable is ln(Total expenditures).

- **a parameter:** ဒီမှာ a တန်ဖိုးက 31.810 ပါ။ ဆိုလိုတာကတော့ ပျမ်းမျှဝင်ငွေ ၁ HUF ရှိတဲ့ Csernely က အိမ်ထောင်စုတစ်စုရဲ့ စုစုပေါင်းအသုံးစရိတ်က 31.810 HUF ရှိတယ်လို့ ဆိုလိုတာပါ။
- **b parameter:** b တန်ဖိုးက 0.683 ပါ။ ဆိုလိုတာကတော့ ပျမ်းမျှဝင်ငွေ ၁ ရာခိုင်နှုန်း တိုးလာရင် စုစုပေါင်းအသုံးစရိတ်က ၀.၆၈၃ ရာခိုင်နှုန်း တိုးလာမယ်လို့ ခန့်မှန်းရပါတယ်။

မှတ်ချက် -

- **Power regression: ŷ = a · X^b**
  - a: x=1 ဖြစ်တဲ့အခါ y=a ဖြစ်ပါတယ်။
  - b: x က ၁ ရာခိုင်နှုန်း တိုးလာရင် y က b ရာခိုင်နှုန်း ပြောင်းလဲသွားဖို့ မျှော်လင့်နိုင်ပါတယ်။
- **Compound regression: ŷ = a · b^x**
  - a: ဒါကိုတော့ အသေးစိတ် မဆန်းစစ်ပါဘူး။
  - b: x က ၁ ယူနစ် တိုးလာရင် y က b ဆ တိုးလာဖို့ မျှော်လင့်နိုင်ပါတယ်။

Chapter 10: TIME SERIES ANALYZES (အချိန်နှင့်အမျှ ပြောင်းလဲမှုများကို ဆန်းစစ်ခြင်း)

Example 1: ၁၉၉၄ ခုနှစ်အတွင်းက လူဦးရေ၊ မွေးဖွားနှုန်းနှင့် သေဆုံးနှုန်း အချက်အလက်များကို ဆန်းစစ်ခြင်း

Month	Population (thousand persons)	Number of birth (persons)	Number of death (persons)
1	10 273	10 238	13 888
2	10 270	9 285	12 825
3	10 267	10 105	12 516
4	10 265	9 617	11 753
5	10 262	9 548	12 328
6	10 260	9 717	11 839
7	10 258	9 965	11 848
8	10 257	9 980	11 722
9	10 256	9 844	10 968
10	10 252	9 021	12 542
11	10 249	8 740	11 743
12	10 246	9 538	12 917

Source: HCSO

- ပထမဆုံးအနေနဲ့ ပေးထားတဲ့ အချက်အလက်တွေကို သုံးပြီး **SPSS data set** တစ်ခု အရင်တည်ဆောက်ရပါမယ်။

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	month	Numeric	8	0	1994	None	None	8	Right	Scale
2	population	Numeric	8	0	population (th p)	None	None	8	Right	Scale
3	birth	Numeric	8	0	number of birth (person)	None	None	8	Right	Scale
4	death	Numeric	8	0	number of death (person)	None	None	8	Right	Scale

	month	population	birth	death
1	1	10273	10238	13888
2	2	10270	9285	12825
3	3	10267	10105	12516
4	4	10265	9617	11753
5	5	10262	9548	12328
6	6	10260	9717	11839
7	7	10258	9965	11848
8	8	10257	9980	11722
9	9	10256	9844	10968
10	10	10252	9021	12542
11	11	10249	8740	11743
12	12	10246	9538	12917

- မွေးဖွားနှုန်း (birth) နဲ့ သေဆုံးနှုန်း (death) အတွက် **sequence chart** တစ်ခု ဆွဲကြည့်ကြရအောင်။

- ဒါကိုလုပ်ဖို့ Menu ထဲက **Analyze > Forecasting > Sequence Charts** ကို သွားပါ။

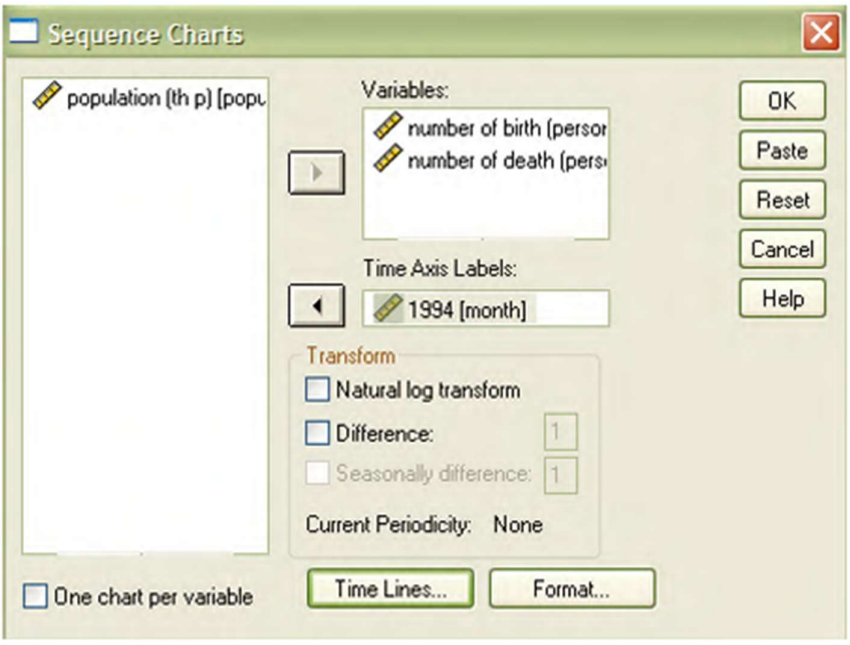


Figure 49 – Sequence Charts Dialog Box

- ပြီးရင် Variables ထဲကို number of birth နဲ့ number of death တို့ကို ထည့်ပြီး၊ Time Axis Labels နေရာမှာတော့ လအလိုက် ပြပေးမယ့် month ကို ရွေးပေးရပါမယ်။

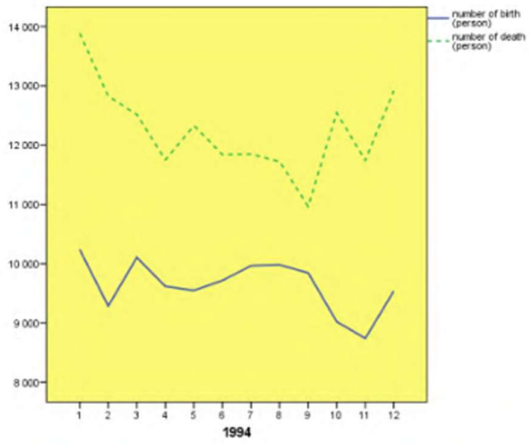
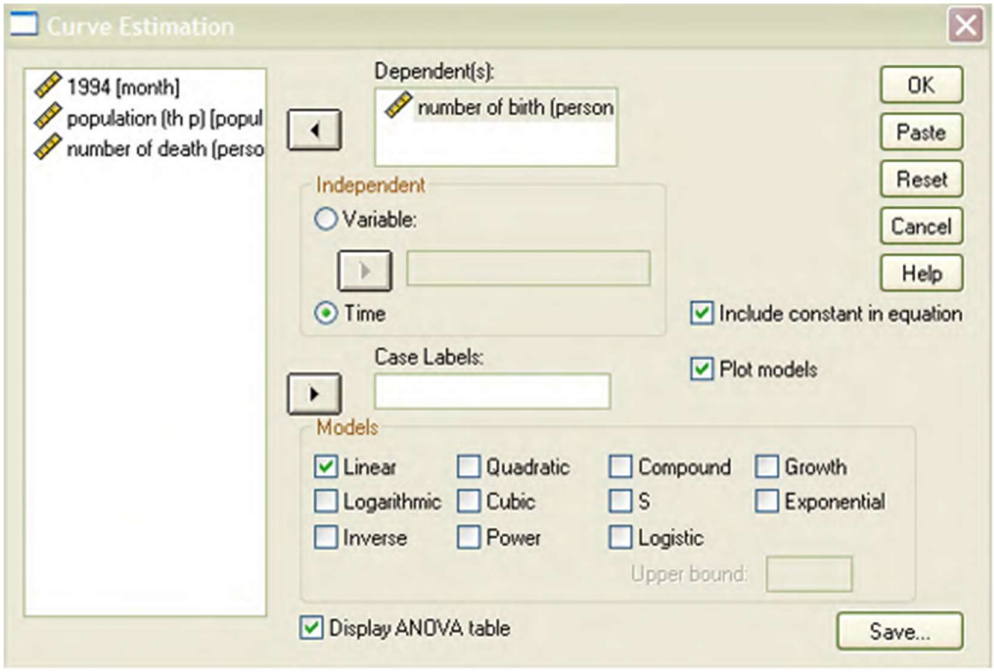


Figure 50 – Sequence Chart

- ဆက်လက်ပြီး မွေးဖွားနှုန်းအတွက် **linear trend model** တစ်ခု တည်ဆောက်ကြည့်ပါမယ်။
- Menu ထဲက **Analyze > Regression > Curve Estimation** ကို သွားပါ။ **Dependent variable** နေရာမှာ number of birth ကို ထည့်ပြီး၊ **Independent variable** နေရာမှာတော့ **Time** ကို ရွေးပေးပါ။



ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	536259,1	1	536259,084	3,261	,101
Residual	1644363	10	164436,258		
Total	2180622	11			

- ရလာတဲ့ **ANOVA** ဇယားအရ significance level က 0.101 ဖြစ်နေတဲ့အတွက် ဒီ linear regression model ဟာ 10.1% significance level မှာမှ ခိုင်မာမှု ရှိတယ်လို့ ပြောလို့ရပါတယ်။

Coefficients

	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.
Case Sequence	-61,238	33,910	-,496	-1,806	,101
(Constant)	10031,212	249,573		40,194	,000

Linear regression line: $\hat{y}=b_0+b_1t=10031.212-61.238t$

- **Coefficients** ဇယားအရ linear regression line ဟာ **$\hat{y}$ = 10031.212 - 61.238t** ဖြစ်ပါတယ်။

- ဒါကို အဓိပ္ပာယ်ဖွင့်ရရင် - ၁၉၉၃ ခုနှစ် ဒီဇင်ဘာလမှာ ပျမ်းမျှ မွေးဖွားနှုန်းဟာ ၁၀,၀၃၁ ယောက် ရှိခဲ့ပြီး၊ လစဉ် ပျမ်းမျှ ၆၂ ယောက်နှုန်း (၆၁.၂၃၈) လျော့နည်းလာနေတယ်လို့ ဆိုလိုတာပါ။

- ၁၉၉၅ ခုနှစ် ဇန်နဝါရီလအတွက် ခန့်မှန်းချက် (estimation) ထုတ်ကြည့်ဖို့အတွက် **Analyze > Regression > Curve Estimation > Save** menu ထဲကို သွားပါ။

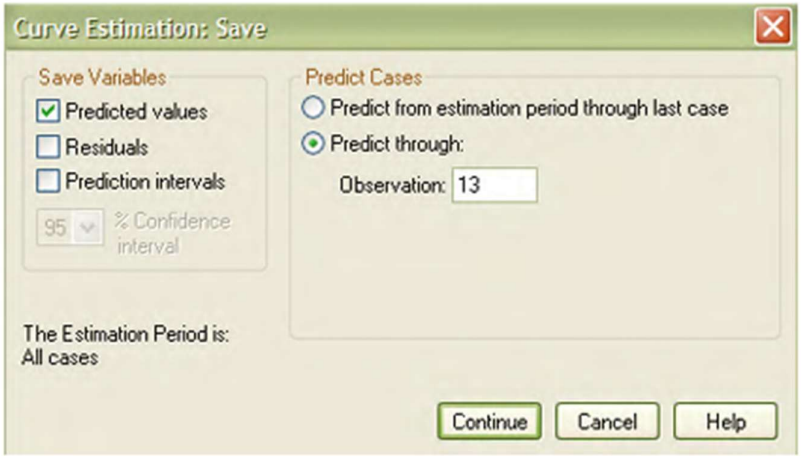


Figure 52 – Curve Estimation: Save Dialog Box

- အဲဒီမှာ **Predict through** ကို ရွေးပြီး၊ ကျွန်တော်တို့ဆီမှာ ၁၂ လစာ ဒေတာ ရှိပြီးသား ဖြစ်တဲ့အတွက် ၁၃ လမြောက်ဖြစ်တဲ့ ဇန်နဝါရီအတွက် **Observation: 13** လို့ ရိုက်ထည့်ပေးရပါမယ်။

- ဒါဆိုရင် **Data Editor** window ထဲက variable အသစ်မှာ ခန့်မှန်းချက် ရလဒ်တွေ ထွက်လာပါလိမ့်မယ်။

	month	population	birth	death	FIT_1
1	1	10273	10238	13888	9969,97436
2	2	10270	9285	12825	9908,73660
3	3	10267	10105	12516	9847,49883
4	4	10265	9617	11753	9786,26107
5	5	10262	9548	12328	9725,02331
6	6	10260	9717	11839	9663,78555
7	7	10258	9965	11848	9602,54779
8	8	10257	9980	11722	9541,31002
9	9	10256	9844	10968	9480,07226
10	10	10252	9021	12542	9418,83450
11	11	10249	8740	11743	9357,59674
12	12	10246	9538	12917	9296,35897
13	.	.	.	.	9235,12121

Figure 53 – Predicted Number of Birth

Example 2: ရထားပို့ဆောင်ရေး အချက်အလက်များမှတစ်ဆင့် Seasonality (ရာသီအလိုက် အပြောင်းအလဲ) ကို လေ့လာခြင်း

Year	Quarters	Railway transport (thousand tons)	Year	Quarters	Railway transport (thousand tons)
1990	1	20 516	1992	3	11 044
1990	2	20 674	1992	4	14 806
1990	3	21 736	1993	1	10 924
1990	4	24 796	1993	2	9 870
1991	1	15 510	1993	3	10 579
1991	2	14 511	1993	4	13 456
1991	3	14 557	1994	1	4 920
1991	4	18 510	1994	2	6 354
1992	1	13 974	1994	3	6 834
1992	2	13 294	1994	4	8 599

- အရင်ဆုံး data တွေကို ရိုက်ထည့်ပြီးရင် **Data > Define Dates...** ကို သွားပြီး အချိန်ကာလတွေကို သတ်မှတ်ပေးရပါမယ်။

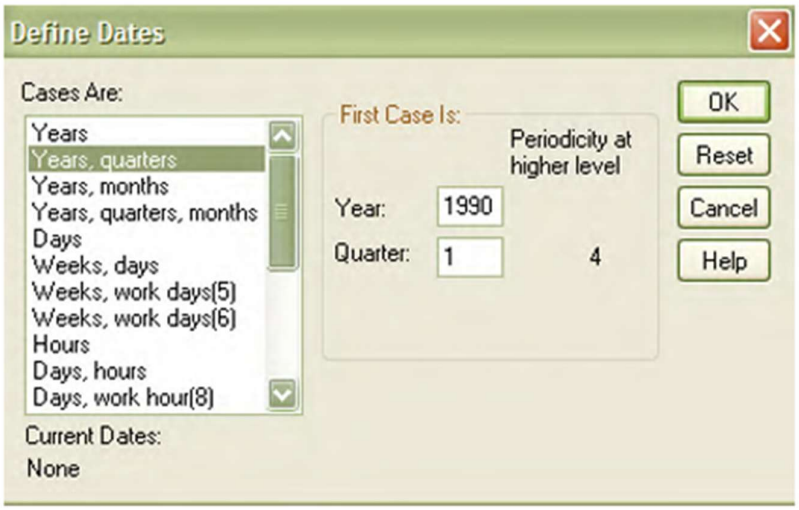


Figure 54 – Define Dates Dialog Box

- ဒီဥပမာမှာတော့ တစ်နှစ်ကို ၄ ကြိမ် ခွဲထားတဲ့အတွက် **Years, quarters** ကို ရွေးပြီး၊ စတင်တဲ့နှစ်နဲ့ quarter ကို သတ်မှတ်ပေးလိုက်ရင် Year, Quarter နဲ့ Date ဆိုတဲ့ column ၃ ခု အသစ် ပေါ်လာပါမယ်။
- **Analyze > Forecasting > Sequence...** ကို သွားပြီး **sequence chart** ဆွဲကြည့်တဲ့အခါမှာ ဒေတာတွေထဲမှာ **seasonality** ပါဝင်နေတာကို တွေ့ရပါလိမ့်မယ်။

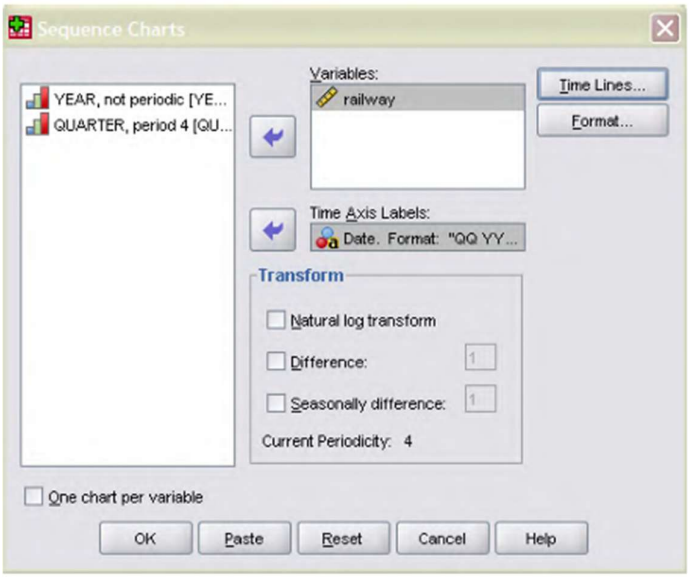


Figure 55 – Sequence Chart Dialog Box

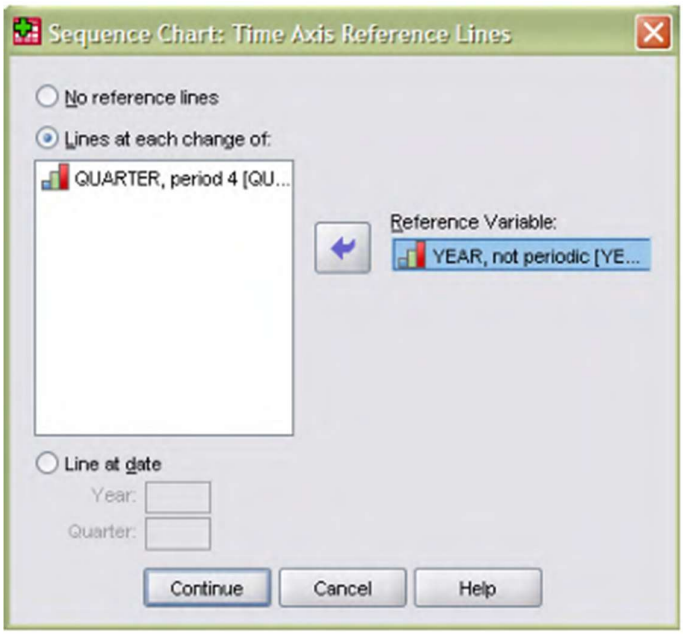


Figure 56 – Time Axis Reference Line Dialog Box

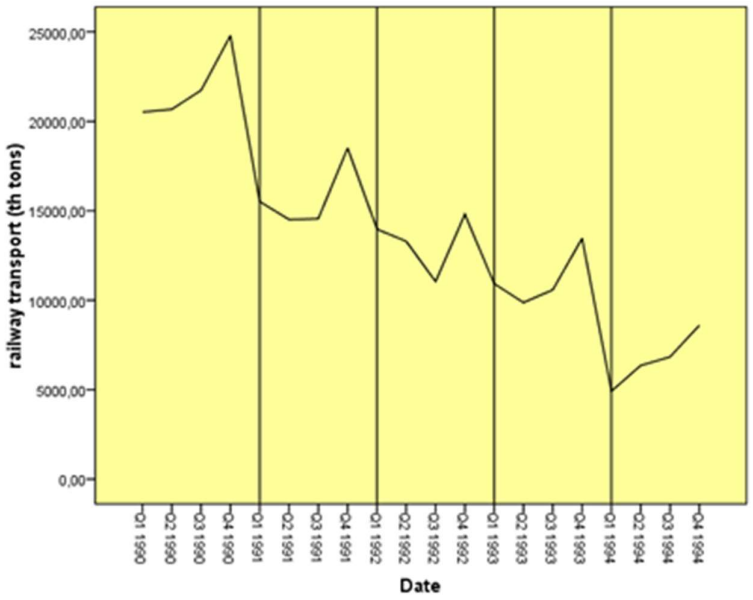


Figure 57 – Time Axis Reference Line Dialog Box

- ဒါကြောင့် **seasonal decomposition** (ရာသီအလိုက် အတက်အကျများကို ခွဲထုတ်ခြင်း) ကို လုပ်ဆောင်ရပါမယ်။

- **Analyze > Forecasting > Seasonal Decomposition** ကို သွားပြီး Variable နေရာမှာ railway transport ကို ထည့်ပါ။

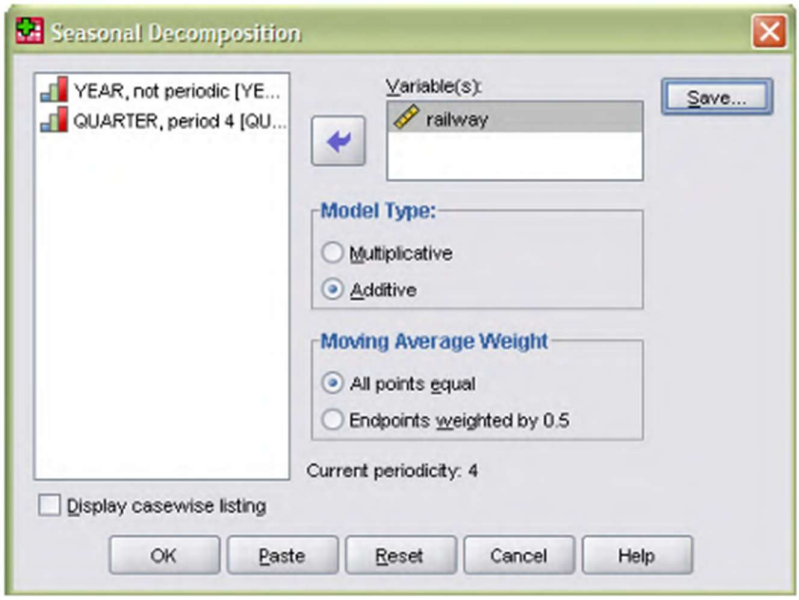


Figure 58 – Seasonal Decomposition Dialog Box

- ဒေတာရဲ့ ပုံသဏ္ဌာန်အရ **Model Type** ကို **Additive** လို့ ရွေးပေးရပါမယ်။
- ဒါဆိုရင် Data Editor ထဲမှာ variable ၄ ခု အသစ် ထပ်တိုးလာပါမယ် -
    - **ERR_1:** error component (အမှားတန်ဖိုး)
    - **SAS_1:** seasonal adjusted series (ရာသီအလိုက် အတက်အကျတွေကို ဖယ်ထုတ်ထားတဲ့ ဒေတာ)
    - **SAF_1:** seasonal factors (ရာသီအလိုက် ပြောင်းလဲစေတဲ့ အချက်များ)
    - **STC_1:** smoothed trend-cycle component (သန့်စင်ပြီးသား trend လမ်းကြောင်း)

For visualization, select Analyze / Forecasting / Sequence Charts

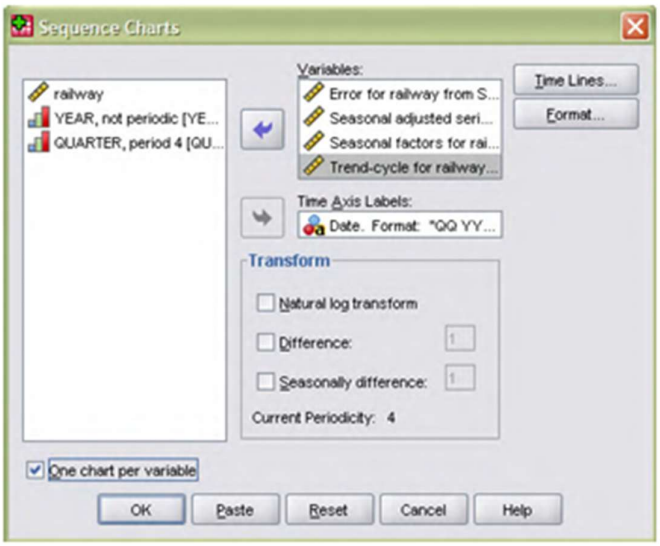


Figure 59 – Seasonal Decomposition Dialog Box

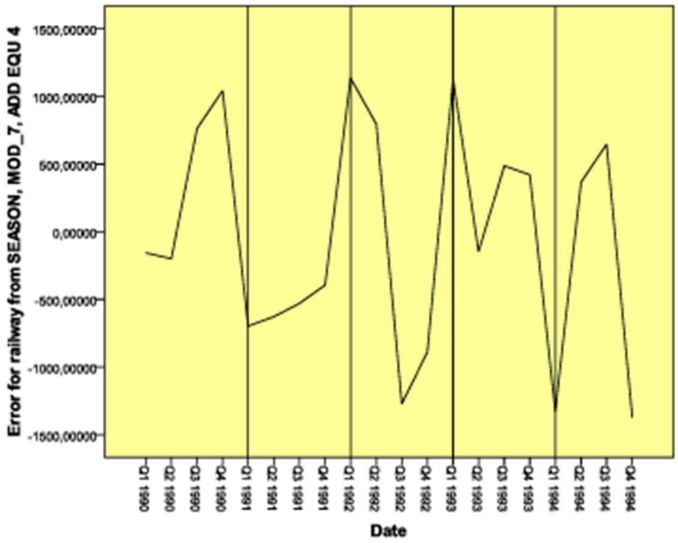


Figure 60 – Error Component

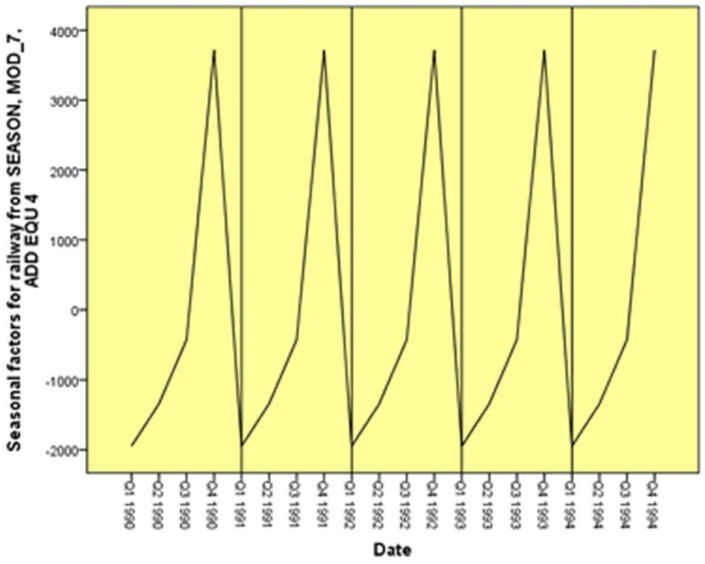


Figure 61 – SAF_1: Seasonal Component

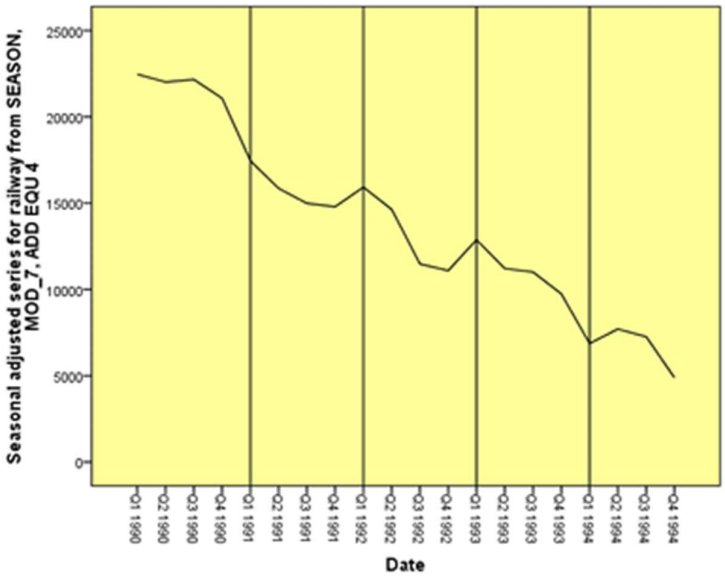


Figure 62 – SAS_1: Component without Seasonality

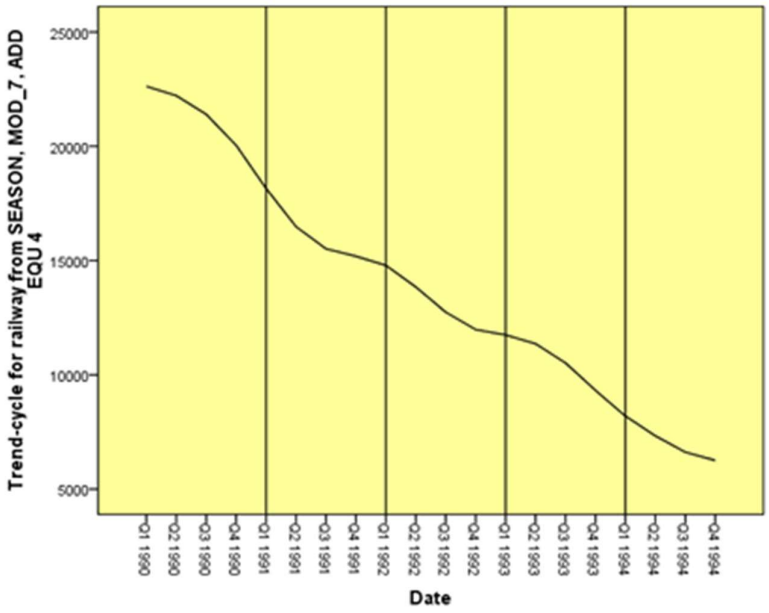


Figure 63 – STC_1: Smoothed Trend-cycle Component

Seasonal Factors

Series Name:railway

Period	Seasonal Factor
1	-1948.49688
2	-1341.87188
3	-426.13438
4	3716.50313

- Output window မှာ တွေ့ရမယ့် **Seasonal Factors** တွေအရ ပထမ သုံးလပတ် ၃ ခုမှာ trend ရဲ့ ခန့်မှန်းချက်က လက်တွေ့ထက် များနေပြီး၊ စတုတ္ထ သုံးလပတ် (SAF_1 = 3716.5) မှာတော့ trend က လက်တွေ့ထက် လျှော့ပြီး ခန့်မှန်းထားတာကို တွေ့ရပါတယ်။
- ၁၉၉၅ ခုနှစ် တတိယ သုံးလပတ် (Quarter III) အတွက် ခန့်မှန်းချက် ထုတ်ချင်ရင် **Analyze > Regression > Curve Estimation > Save** ကို သွားပြီး ၁၉၉၅ Quarter 3 အထိ ခန့်မှန်းခိုင်းရပါမယ်။

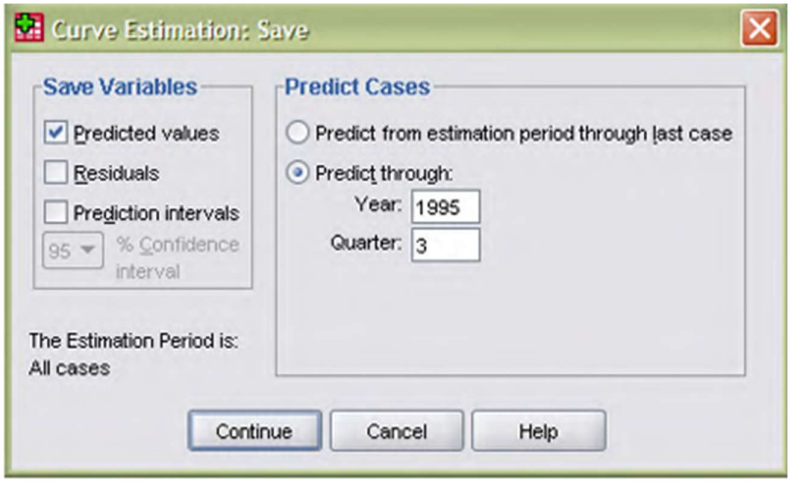


Figure 64 – Forecasted Values

	railway	YEAR_	QUARTER_	DATE_	ERR_1	SAS_1	SAF_1	STC_1	FIT_2
1	20516,00	1990	1	Q1 1990	-156,74722	22464,49688	-1948,49688	22621,24410	21482,45714
2	20674,00	1990	2	Q2 1990	-198,29583	22015,87188	-1341,87188	22214,16771	20670,95639
3	21736,00	1990	3	Q3 1990	762,11944	22162,13438	-426,13438	21400,01493	19859,45564
4	24796,00	1990	4	Q4 1990	1040,77500	21079,49688	3716,50313	20038,72188	19047,95489
5	15510,00	1991	1	Q1 1991	-695,44722	17458,49688	-1948,49688	18153,94410	18236,45414
6	14511,00	1991	2	Q2 1991	-626,55833	15852,87188	-1341,87188	16479,43021	17424,95338
7	14557,00	1991	3	Q3 1991	-530,54722	14963,13438	-426,13438	15513,68160	16613,45263
8	18510,00	1991	4	Q4 1991	-393,22500	14793,49688	3716,50313	15186,72188	15801,95188
9	13974,00	1992	1	Q1 1992	1135,88611	15922,49688	-1948,49688	14786,61076	14990,45113
10	13294,00	1992	2	Q2 1992	794,10833	14635,87188	-1341,87188	13841,76354	14178,95038
11	11044,00	1992	3	Q3 1992	-1269,43611	11470,13438	-426,13438	12739,57049	13367,44962
12	14806,00	1992	4	Q4 1992	-888,44722	11089,49688	3716,50313	11977,94410	12555,94887
13	10924,00	1993	1	Q1 1993	1128,55278	12872,49688	-1948,49688	11743,94410	11744,44812
14	9870,00	1993	2	Q2 1993	-145,89167	11211,87188	-1341,87188	11357,76354	10932,94737
15	10579,00	1993	3	Q3 1993	487,45278	11005,13438	-426,13438	10517,68160	10121,44662
16	13456,00	1993	4	Q4 1993	420,21944	9739,49688	3716,50313	9319,27743	9309,94586
17	4920,00	1994	1	Q1 1994	-1325,00278	6868,49688	-1948,49688	8193,49965	8498,44511
18	6354,00	1994	2	Q2 1994	366,21944	7695,87188	-1341,87188	7329,65243	7686,94436
19	6834,00	1994	3	Q3 1994	647,30000	7260,13438	-426,13438	6612,83437	6875,44361
20	8599,00	1994	4	Q4 1994	-1371,92847	4882,49688	3716,50313	6254,42535	6063,94286
21	.	1995	1	Q1 1995	.	.	.	.	5252,44211
22	.	1995	2	Q2 1995	.	.	.	.	4440,94135
23	.	1995	3	Q3 1995	.	.	.	.	3629,44060

Figure 65 – Predicted Railway Transport

- ရလာတဲ့ **FIT_2** (ခန့်မှန်းချက်) တန်ဖိုးထဲကနေ သက်ဆိုင်ရာ **seasonal factor** ကို ပြန်ပေါင်း/နှုတ်ပြီးမှ တကယ့် ခန့်မှန်းချက် အစစ်အမှန်ကို ရရှိမှာ ဖြစ်ပါတယ်။